



Teoría de la Relatividad General

Universidad de Granada, 4° **curso Física**

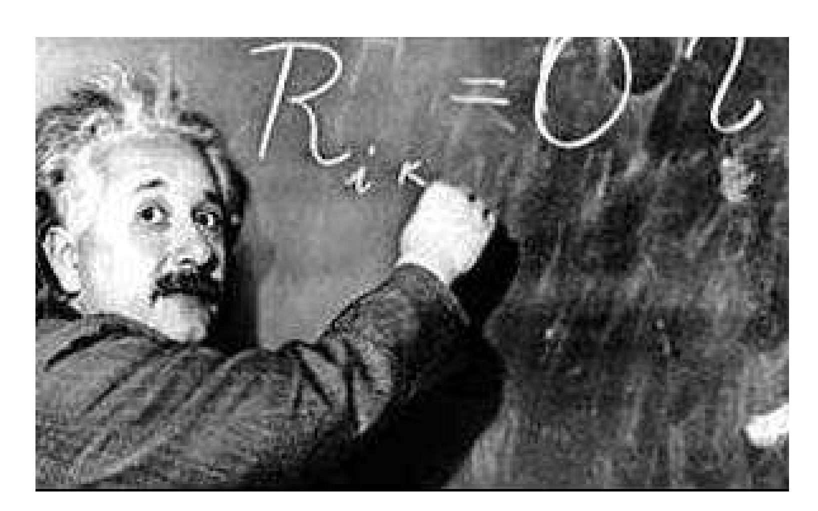
22 de septiembre de 2013

Bert Janssen

Dpto de Física Teórica y del Cosmos,
Edificio Mecenas, Campus de Fuente Nueva
Universidad de Granada, 18071 Granada, España
`bjanssen@ugr.es`



Relatividad por M.C. Escher (1953)



"I want to know God's thoughts. The rest are details."

A. Einstein.

○ Ulm, Alemania, 14 marzo 1879
† Princeton, Estados Unidos, 18 abril 1955

Comentario sobre el dibujo "*Relatividad*" de M.C. Escher (1902 - 1972):
Tres mundos completamente distintos están construidos en una sola unidad inseparable. [...] *Podemos dividir las 16 figuritas que aparecen en el dibujo en tres grupos, cada uno de los cuales vive en su propio mundo. Para cada grupo todo lo que aparece en el dibujo es parte de su mundo, sólo ven las cosas de distinta manera y le dan nombres distintos. Lo que para un grupo es un techo, es una pared para el otro. Lo que en una comunidad es una puerta, considera la otra como una escotilla en el suelo.*
(Bruno Ernst, *El espejo Mágico*, 1978)

Derechos de autor

Al ser un libro de texto, este libro no pretende ser un trabajo original, sino claramente está basado en gran parte en otros libros de texto y artículos originales. Para hacer la lectura más amena, hemos omitido las referencias explícitas, pero se pueden encontrar todas las obras consultadas en el Appendice C.

Para cualquier duda, contacte con el autor en `bjanssen@ugr.es`.
Versiones actualizadas de este libro se pueden encontrar en
`http://www.ugr.es/~bjanssen/text/BertJanssen-RelatividadGeneral.pdf`

Granada, 4 febrero 2013

Copyright

Being a textbook, this works pretends by no means to be original, but clearly relies heavily on other textbooks and original articles. In order to make the text more readable, we have ommitted the explicite references in the text, but a complete list of consulted works can be found in the Appendix C.

For any questions, contact the author at `bjanssen@ugr.es`.
Updated versions of this book can be found at
`http://www.ugr.es/~bjanssen/text/BertJanssen-RelatividadGeneral.pdf`

Granada, 4 february 2013

Índice general

Prefacio

Las cosas se deberían explicar de la forma más sencilla posible, pero no más sencilla aún.
(A. Einstein)

Alrededor de 1920 un peridodista preguntó al astrofísico británico Arthur Eddington si era verdad que en el mundo entero sólo había cuatro personas que entendían la teoría de la relatividad general. Eddington se quedó pensativo durante unos momentos y respondió: "Me estoy preguntando quién podría ser el cuarto."

Aunque es posible que la anécdota sea históricamente correcta, la afirmación de Eddington ciertamente no lo era, ni ahora, ni entonces. Prueba de ello es la gran cantidad de físicos que se han puesto a trabajar en la relatividad general y el número de soluciones que fueron halladas en los años inmediatamente después de la publicación de la teoría en 1915: la acción de Hilbert (1915), los agujeros negros de Schwarzschild (1916) y Reissner-Nordström (1916 y 1918), los espacios de De Sitter y anti-De Sitter (1916 y 1917), la clasificación cosmológica de Friedmann (1922), la onda gravitacional de Brinkmann (1923), la compactificaciones de Kaluza y Klein (1921 y 1926), ...

Aún así la teoría de la relatividad siempre ha tenido fama de contraintuitiva y tremendamente dificil. Esta fama, en mi opinión, es solamente justificada hasta cierto punto. La teoría de la relatividad especial forma parte del curriculum estándar de cualquier estudiante de física en los primeros años de la carrera. La relatividad general es más dificil y la suelen estudiar sólo los estudiantes especializándose en física teórica.

Sin embargo, conviene distinguir dos partes dentro de la relatividad general: una parte física, que describe las ideas básicas de la teoría, y una parte matemática, que nos da el formalismo con el cual describir la física. La dificultad de la relatividad general está en la parte matemática, ya que contiene análisis tensorial y geometría diferencial, temas que el típico estudiante de física no encuentra en su curriculum sino para estudiar relatividad general. Por otro lado, la parte física es relativamente sencilla, no siendo en el fondo nada más que llevar hasta sus últimas consecuencias lógicas unos pocos principios básicos. Según Rutherford "una buena teoría física se le puede explicar a una camarera en un bar". En mi opinión (la parte física de) la relatividad general satisface esta condición (con suficiente tiempo disponible).

En este curso he intentado mantener el equilibrio entre estas dos partes, puesto que creo que las dos son igual de importantes: sin la matemática es imposible alcanzar un nivel que vaya más allá de los numerosos libros de popularización sobre relatividad. Por otro lado, sin la parte física uno estaría haciendo básicamente geometría diferencial a un nivel que lo podía haber hecho Riemann en el siglo XIX y se perdería unos profundos principios sobre la equivalencia de observadores y la estructura del espacio y el tiempo.

La teoría de la relatividad es, junto con la mecánica cuántica, uno de los dos pilares de la física del siglo XX. Cualquier teoría que pretenda ser fundamental debería tomar en cuenta estos dos pilares. Pero hay dos razones que hacen que la teoría de la relatividad sea históricamente diferente de la mecánica cúantica, incluso de la teoría cúantica de campos.

La primera razón es que la teoría de la relatividad fue esencialmente el trabajo de una sola persona. Donde la mecánica cúantica fue desarrollado por gente como Bohr, de Broglie, Heisenberg, Schrödinger, Pauli, etc, en partes más o menos igual de importantes, la teoría de la relatividad es el fruto (casi) únicamente de Albert Einstein, además en un periodo de unos 10 años, entre 1905 y 1916.

La segunda razón que hace que la teoría de la relatividad ocupe un lugar especial en la física es que fue desarollada por razones puramente teóricas, no para dar una explicación a unos

resultados experimentales nuevos. Muchas veces se dice que la relatividad especial fue desarollada para explicar el resultado nulo de los experimentos de Michelson y Morley, pero esto no es históricamente correcto. Einstein mismo siempre ha dicho que en 1905 no sabía de los resultados de Michelson.[1]

El verdadero punto de partida de Einstein era la incompatibilidad de la mecánica newtoniana, la teoría de Maxwell y el principio de la relatividad de Galilei. La mecánica newtoniana y la teoría de Maxwell tienen grupos de simetría diferentes, mientras el principio de la relatividad dice en grandes líneas que todas las teoría físicas deberían tener el mismo grupo. La solución de Einstein a este problema teórico, la teoría de la relatividad especial, es una reformulación de la mecánica newtoniana en términos del grupo de Lorentz, el grupo de simetría de la teoría de Maxwell. De paso la relatividad especial nos proprociona una nueva manera de pensar sobre la estructura del espacio y el tiempo.

También la motivación para la relatividad general fue puramente teórica: Einstein se dió cuenta de que la teoría de la gravedad, tal como fue propuesta por Newton, no era compatible con la estructura del espacio y el tiempo que surge de la relatividad especial. Einstein formuló una nueva versión de la gravedad, que esencialmente convierte la gravedad newtoniana en una teoría de campos, un concepto introducido por Faraday y Maxwell unos 50 años antes. La interacción gravitatoria ya no es instantánea y a distancia, sino a través de un campo intermediario por el cual la fuerza gravitatoria se propaga con velocidad finita. Lo revolucionario de la relatividad general es la identificación de este campo intermediario con la métrica, un objeto matemático que describe las propiedades geométricas del espacio. La relatividad general induce por lo tanto una profunda relación entre la gravedad y la curvatura del espaciotiempo.

De este modo la teoría de la relatividad no es sólo una teoría moderna de la gravedad, mejorando la gravedad newtoniana, sino también nos enseña unas lecciones en la frontera entre física y metafísica. Primero, por un lado la relatividad especial ha eliminado los conceptos del espacio absoluto, del tiempo absoluto y de la velocidad absoluta, por no ser observables, mientras por otro lado la relatividad general ha incorporado en la física el concepto del espaciotiempo dinámico, como una entidad física, igualmente real que conceptos como masa, carga, energía o momento angular. El espaciotiempo ha pasado de ser un escenario estático donde ocurre la física a ser una parte más de la física que influye lo que contiene y puede ser influenciado por ello.

Y la segunda lección que nos enseña la teoría de la relatividad es que una buena teoría fisica tiene que hacer algo más que simplemente reproducir las observaciones o los datos experimentales de un observador. Debería poder escribir los datos experimentales de cualquier observador y si distintos observadores están relacionados por ciertas transformaciones de simetría, entonces la teoría debería reflejar estas simetrías y tomar una forma tal que es invariante bajo estas transformaciones. En otras palabras, la teoría de la relatividad nos enseña la forma en que de debemos formular una teoría física para poder tomarla en serio.

No es de extrañar que la teoría de la relatividad sea una de los pilares fundamentales de la física conocida.

[1] Aunque esto tampoco es verdad: existen pruebas de que lo había discutido en varias ocasiones en su grupo de amigos físicos.

Parte I

El Principio de la Relatividad y la Relatividad Especial

Capítulo 1

Breve repaso de la teoría de Maxwell

La teoría de la relatividad especial debe su origen a las ecuaciones de Maxwell del campo electromagnético.
(A. Einstein)

La teoría de Maxwell de las interacciones electromagnéticas siempre ha tenido una relación muy estrecha con la teoría de la relatividad. Aunque data de unos 40 años antes de que Einstein presentara su relatividad especial, ya llevaba algunas de las características de ésta, como la velocidad de la luz como una velocidad absoluta y la invariancia bajo las transformaciones de Lorentz. La teoría de Maxwell encaja tan extraordinariamente bien en la relatividad especial y con unas ligeras modificaciones en la relatividad general, que servirá en numerosas ocasiones como ejemplo concreto o caso de prueba de técnicas que encontramos a lo largo de este curso. Por lo tanto merece la pena repasar brevemente los aspectos más importantes de la teoría de Maxwell en lenguaje tridimensional.

1.1. El límite a sistemas continuos

La teoría de Maxwell es una *teoría de campos*, es decir una teoría con infinitos grados de libertad. El concepto de un campo físico fue introducido por Michael Faraday (1791-1869), que había observado que el serrín metálico se agrupaba según filamentos en la proximidad de cargas eléctricas e imanes. Faraday se imaginaba que las cargas y los imanes generaban campos eléctricos y magnéticos que se extendían por el espacio y que éstos actuaban a su vez sobre otras cargas e imanes. De esta manera, Faraday consiguió eliminar el problema de la acción a distancia, tan típico de la gravedad newtoniana y la ley de Coulomb, que postula que las interacciones físicas se manifiestan instantáneamente a distancias arbitrariamente grandes, sin preocuparse de la pregunta de cómo se propagan estas influencias. En una teoría de campos, como la teoría de Maxwell, las distintas partículas no interacciones direcamente entre ellas, sino usan los campos como los transportadores de las interacciones físicas: una perturbación de una carga eléctrica genera una perturbación en el campo electromagnético, que a su vez se propaga a velocidad finita (la velocidad de la luz) por todo el espacio. Las demás cargas en el universo sólo notan la influencia de la perturbación original, cuando la perturbación del campo electromagnético haya llegado hasta ellas. En este sentido un campo físico es realmente un intermediario a través del cual las distintas partículas interaccionan sobre grandes distancias. Hoy en día, las teorías de campos forman la base de la física moderna: toda la física de partículas está basada en la teoría cuántica de campos, mientras en cierto modo la relatividad general no es más que una reformulación de la gravedad newtoniana en lenguaje de una teoría de campos.

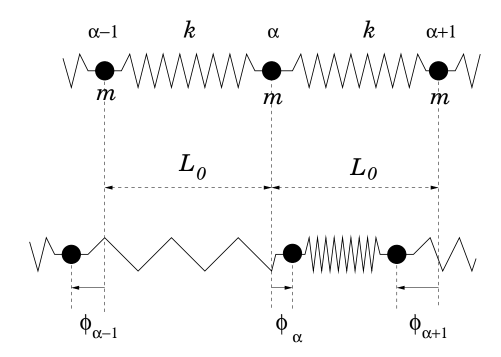


Figura 1.1: *Una serie de masas m están conectadas a través de muelles con constante k y posición de equilibrio L_0. En la situación de equilibrio (arriba) las masas están separadas por la distancia L_0. En la situación general (abajo) el desplazamiento de la masa m de la posición de equilibrio está caracterizado por ϕ_α.*

Conceptualmente las teorías de campos son un poco diferentes que los sistemas discretos. Por lo tanto es útil estudiar la conexión con la mecánica analítica discreta, antes de tratar a fondo el electromagnetismo. Desde el punto de vista mecánico, una teoría de campos no es nada más que una teoría con un número infinito (continuo) de grados de libertad. Son aplicables por lo tanto las mismas técnicas que ya conocemos de (por ejemplo) el formalismo lagrangiano, sólamente tomando en cuenta la sutileza de tomar de manera adecuada el límite continuo. En esta sección revisaremos cómo tomar este límite.

Consideramos un sistema que consiste de una serie infinita (pero contable) de masas iguales alineadas a lo largo del eje x y conectadas por muelles idénticos de tamaño L_0 y constante elástica k. Supondremos además que las masas sólo se pueden mover en la dirección x (Véase Figura 1.1). Tomamos como coordenadas generalizadas $q_\alpha(t)$ la posición de la α-ésima masa m

$$q_\alpha(t) = \alpha L_0 + \phi_\alpha(t), \tag{1.1}$$

donde el indice $\alpha \in \mathbb{Z}$ corre de $-\infty$ a ∞ y $\phi_\alpha(t)$ mide la desviación de la posición de equilibrio. Las velocidadas generalizadas por lo tanto vienen dadas por $\dot{q}_\alpha = \dot{\phi}_\alpha$. La energía potencial es proporcional al cuadrado de la desviación de los muelles del tamaño de equilibrio y viene dada por

$$V = \frac{1}{2}\sum_\alpha k(q_{\alpha+1} - q_\alpha - L_0)^2 = \frac{1}{2}\sum_\alpha k(\phi_{\alpha+1} - \phi_\alpha)^2. \tag{1.2}$$

Podemos por lo tanto escribir el lagrangiano como

$$\begin{aligned} L &= \frac{1}{2}\sum_\alpha m\dot{\phi}_\alpha^2 - \frac{1}{2}\sum_\alpha k(\phi_{\alpha+1} - \phi_\alpha)^2 \\ &= \frac{1}{2}\sum_\alpha L_0\Big[\frac{m}{L_0}\,\dot{\phi}_\alpha^2 - kL_0\Big(\frac{\phi_{\alpha+1} - \phi_\alpha}{L_0}\Big)^2\Big], \end{aligned} \tag{1.3}$$

donde en la última igualdad hemos sacado un factor L_0 por razones que se harán claras un poco más adelante. Las ecuaciones de movimiento vienen dadas por

$$\frac{m}{L_0}\ddot{\phi}_\alpha = kL_0\Big(\frac{\phi_{\alpha+1} - \phi_\alpha}{L_0^2}\Big) - kL_0\Big(\frac{\phi_\alpha - \phi_{\alpha-1}}{L_0^2}\Big). \tag{1.4}$$

Aquí no estamos interesados en intentar resolver estas ecuaciones, sino queremos saber qué ocurre con el lagrangiano y la ecuaciones de movimiento en el límite donde la posición de equilibrio L_0 tiende a cero.

Una primera observación es que en este caso la distribución de masas se hace continua y el sistema de masas conectadas por muelles tiende a una varilla elástica con cierta densidad de masa ρ. El índice discreto α que hace referencia a la α-ésima masa, se convierte en una variable continua x, indicando los puntos a lo largo de la varilla. Concretamente, la coordenada generalizada $\phi_\alpha(t)$ que indica el desplazamiento de la masa α se convierte en una función $\phi(x,t)$ del desplazamiento de una cantidad infinitesimal de masa dm a lo largo de la recta.

De este modo hemos pasado de una teoría con un número discreto de grados de libertad a una teoría con un número continuo de grados de libertad. En otras palabras, a una teoría de campos. Nótese que, a pesar de que $\phi(x,t)$ es una función de x y de t sin que ninguna de los dos variables juegue un rol más fundamental que el otro, el origen de cada uno de las variables es muy distinto. Merece la pena enfatizar que en el límite de teoría de campos, las cantidades físicas fundamentales, las que juegan el papel de coordenada generalizada, son los campos $\phi(x,t)$ y no las posiciones x. Las posiciones x no son en cierto modo más que una etiqueta continua para distinguir los diferentes grados de libertad ϕ.

Miremos ahora cómo se transforman el lagrangiano (1.3) y las ecuaciones de movimiento (1.4) en el límite $L_0 \to 0$. Está claro que en este caso las cantidades $\dot{\phi}_\alpha(t)$ y $\ddot{\phi}_\alpha(t)$ se convierten en derivadas parciales con respecto al tiempo:

$$\dot{\phi}_\alpha(t) \longrightarrow \frac{\partial}{\partial t}\phi(x,t), \qquad\qquad \ddot{\phi}_\alpha(t) \longrightarrow \frac{\partial^2}{\partial t^2}\phi(x,t). \tag{1.5}$$

Además, el segundo término de (1.3) es prácticamente por definición la derivada parcial con respecto a x (ésta es la razón por haber sacado el factor L_0 antes),

$$\lim_{L_0\to 0} \frac{\phi_{\alpha+1}(t)-\phi_\alpha(t)}{L_0} \;=\; \lim_{L_0\to 0}\frac{\phi(x+L_0,t)-\phi(x,t)}{L_0} \;=\; \frac{\partial}{\partial x}\phi(x,t) \tag{1.6}$$

y consecuentemente los términos del lado derecho de (1.4) se convierten en una derivada segunda:

$$\lim_{L_0\to 0}\Big[\Big(\frac{\phi_{\alpha+1}-\phi_\alpha}{L_0^2}\Big)-\Big(\frac{\phi_\alpha-\phi_{\alpha-1}}{L_0^2}\Big)\Big] = \lim_{L_0\to 0}\frac{1}{L_0}\Big[\frac{\partial\phi(x+L_0,t)}{\partial x}-\frac{\partial\phi(x,t)}{\partial x}\Big] = \frac{\partial^2\phi(x,t)}{\partial x^2}. \tag{1.7}$$

Queda por lo tanto identificar el límite de los parámetros que aparecen en el lagrangiano y la ecuación de movimiento. La cantidad m/L_0 tiene dimensión $\mathbf{ML}^{-1}$,[1] y en el límite continuo se convierte en la densidad ρ de la varilla. La cantidad kL_0 se identifica de la siguiente manera: la fuerza necesaria para extender uno de los muelles la distancia $(\phi_{\alpha+1}-\phi_\alpha)$ viene dada por

$$F \;=\; k(\phi_{\alpha+1}-\phi_\alpha) \;=\; kL_0\,\frac{(\phi_{\alpha+1}-\phi_\alpha)}{L_0} \;=\; Y\Delta L_0, \tag{1.8}$$

donde $\Delta L_0 = (\phi_{\alpha+1}-\phi_\alpha)/L_0$ es la extensión por unidad de longitud y Y es el módulo de Young, la constante de proporcionalidad entre la fuerza aplicada y la extensión por unidad de longitud en los materiales elásticos, una constante que depende del material considerado.

Con esta identificación de la densidad y el módulo de Young podemos escribir el lagrangiano (1.3) como

$$L = \int dx\Big[\frac{1}{2}\rho\,(\partial_t\phi)^2 - \frac{1}{2}Y(\partial_x\phi)^2\Big], \tag{1.9}$$

donde la suma sobre los desplazamientos $\sum_\alpha L_0$ se ha convertido en una integral $\int dx$ sobre la variable continua x. Del mismo modo, la ecuación de movimiento (1.4) se convierte en una ecuación de ondas

$$\rho\,\partial_t^2\phi(x,t) \;-\; Y\,\partial_x^2\phi(x,t) = 0, \tag{1.10}$$

[1]Para el análisis dimensional usado en este libro referimos al Appendice **??**.

que describe ondas (longitudinales) de densidad en el material de la varilla que se propagan con una velocidad $v = \sqrt{Y/\rho}$.

Podemos escribir esta integral como $L = \int dx\, \mathcal{L}$, donde el integrando $\mathcal{L}(\phi, \partial_x\phi, \partial_t\phi)$ es la *densidad lagrangiana*, que contiene toda la física del sistema. A su vez se define la *acción* S como

$$S(\phi, \partial_x\phi, \partial_t\phi) = \int dt\; L(\phi, \partial_x\phi, \partial_t\phi) = \int dtdx\; \mathcal{L}(\phi, \partial_x\phi, \partial_t\phi). \tag{1.11}$$

Nótese que hemos tenido mucho cuidado en obtener el límite continuo de tanto el lagrangiano (1.10) como las ecuaciones de movimiento (1.9), sin decir todavía nada sobre cómo derivar las ecuaciones de movimiento del lagrangiano. En principio no es muy diferente a la derivación estándar de la mecánica analítica con variables discretas, salvo que ahora hay que variar con respecto a los campos $\phi(x,t)$, es decir, a variables continuas. La herramienta matemática necesaria para esta operación es la *derivada funcional*, definida como

$$\frac{\delta\phi(x',t')}{\delta\phi(x,t)} = \delta(x-x')\delta(t-t'), \qquad\qquad \frac{\delta F(\phi(x',t'))}{\delta\phi(x,t)} = \frac{\partial F(\phi(x',t'))}{\partial\phi(x',t')}\frac{\delta\phi(x',t')}{\delta\phi(x,t)}, \tag{1.12}$$

para cualquier función $F(\phi(x,t))$ y donde $\delta(x-x')$ es la delta de Dirac.

La variación de (1.11) con respecto a los campos $\phi(x,t)$ viene entonces dada por

$$\begin{aligned} 0 \;\equiv\; \delta S = \int dtdx \quad \Big[& \frac{\delta\mathcal{L}(\phi(x,t), \partial_x\phi(x,t), \partial_t\phi(x,t))}{\delta\phi(x',t')}\delta\phi(x',t') \\ & + \frac{\delta\mathcal{L}(\phi(x,t), \partial_x\phi(x,t), \partial_t\phi(x,t))}{\delta\partial_{x'}\phi(x',t')}\delta\partial_{x'}\phi(x',t') \\ & + \frac{\delta\mathcal{L}(\phi(x,t), \partial_x\phi(x,t), \partial_t\phi(x,t))}{\delta\partial_{t'}\phi(x',t')}\delta\partial_{t'}\phi(x',t') \Big]. \end{aligned} \tag{1.13}$$

Igual que en el caso de variables discretas, podemos suponer que $\delta\partial_{x'}\phi(x',t') = \partial_{x'}\delta\phi(x',t')$, de modo que integrando por partes los últimos dos términos e imponiendo las condiciones de contorno

$$\delta\phi(x,t_1) \;=\; \delta\phi(x,t_2) \;=\; \delta\phi(x_1,t) \;=\; \delta\phi(x_2,t) \;=\; 0, \tag{1.14}$$

(es decir, tomando la variación igual a cero tanto en los puntos iniciales y finales como en los contornos[2]), tenemos que la variación toma la forma

$$0 \;\equiv\; \delta S = \int dtdx\; \Big[\frac{\partial}{\partial t}\Big(\frac{\delta\mathcal{L}}{\delta(\partial_t\phi(x,t))}\Big) + \frac{\partial}{\partial x}\Big(\frac{\delta\mathcal{L}}{\delta(\partial_x\phi(x,t))}\Big) - \frac{\delta\mathcal{L}}{\delta\phi(x,t)}\Big]\delta\phi(x,t), \tag{1.15}$$

lo que por el cálculo variacional sólo es cero si está satisfecha la ecuación de Euler-Lagrange para una teoría de campos:

$$\frac{\partial}{\partial t}\Big(\frac{\delta\mathcal{L}}{\delta(\partial_t\phi)}\Big) + \frac{\partial}{\partial x}\Big(\frac{\delta\mathcal{L}}{\delta(\partial_x\phi)}\Big) - \frac{\delta\mathcal{L}}{\delta\phi} = 0. \tag{1.16}$$

Concretamente para el caso de la acción (1.9) tenemos que

$$\frac{\delta\mathcal{L}\Big(\phi(x',t'), \partial_{x'}\phi(x',t'), \partial_{t'}\phi(x',t')\Big)}{\delta(\partial_t\phi(x,t))} = \rho\; \partial_{t'}\phi(x',t')\delta(x-x')\delta(t-t') = \rho\; \partial_t\phi(x,t), \tag{1.17}$$

$$\frac{\delta\mathcal{L}\Big(\phi(x',t'), \partial_{x'}\phi(x',t'), \partial_{t'}\phi(x',t')\Big)}{\delta(\partial_{x'}\phi(x,t))} = Y\; \partial_x\phi(x',t')\delta(x-x')\delta(t-t') = Y\; \partial_x\phi(x,t), \tag{1.18}$$

[2]En el caso de la varilla elástica hemos tomado $x_1 = -\infty$ y $x_2 = \infty$, pero en general la integración se puede hacer tanto en intervalos finitos como infinitos.

tal que (1.16) aplicado al lagrangiano (1.9) implica la ecuación de movimiento (1.10), como hemos derivado antes calculando el límite explícitamente.

En la práctica, las derivadas funcionales, por muy complicadas que parezcan, se aplican como si fueran derivadas ordinarias, olvidándose de que uno en realidad está derivando con respecto a funciones. Omitiendo la dependencia funcional, podemos por lo tanto reducir (1.17) y (1.18) a

$$\frac{\delta\mathcal{L}}{\delta(\partial_t\phi)} = \rho\,\partial_t\phi, \qquad\qquad \frac{\delta\mathcal{L}}{\delta(\partial_x\phi)} = Y\,\partial_x\phi. \tag{1.19}$$

La generalización de todo el formalismo a dos y tres dimensiones debería ser obvia. En lugar de ser funciones de sólo x y t, los campos $\phi(x^i, t)$ van a depender en general de x^i y t y el lagrangiano será una función de $\phi(x^i, t)$ y sus derivadas $\partial_j\phi(x^i, t)$ y $\partial_t\phi(x^i, t)$. La ecuación de Lagrange (1.16) será por lo tanto de la forma

$$\frac{\partial}{\partial t}\Big(\frac{\delta\mathcal{L}}{\delta(\partial_t\phi)}\Big) + \sum_i \frac{\partial}{\partial x^i}\Big(\frac{\delta\mathcal{L}}{\delta(\partial_i\phi)}\Big) - \frac{\delta\mathcal{L}}{\delta\phi} = 0. \tag{1.20}$$

1.2. Las leyes de Maxwell

El físico escocés James Clerk Maxwell (1831-1879) publicó sus cuatro leyes de Maxwell en 1865, aunque la mayoría de ellas ya habían sido descubiertos por Charles-Augustin Coulomb (1736-1806), Hans Christan Ørsted (1777-1851), André-Marie Ampère (1775-1836), Jean-Baptiste Biot (1774-1862), Félix Savart (1791-1841) y Faraday (1791-1867) a base de investigación experimental. De hecho Maxwell añadió sólo un término nuevo a las ecuaciones que ahora llevan su nombre. Pero el gran logro de Maxwell fue unificar el conjunto de leyes empíricas sueltas sobre electrostática, corrientes eléctricas e inducción magnética en una sólida teoría que describe todos los fenómenos relacionados con el electromagnetismo. Y como extra resultó que su teoría era capaz de dar un fundamento teórico a la óptica, una parte de la física que hasta entonces parecía completamente disconexa de los fenómenos electromagnéticos.

Dada una densidad de cargas eléctricas $\rho(\vec{x}, t)$ y una densidad de corriente $\vec{\jmath}\,(\vec{x}, t)$, las leyes de Maxwell para los campos eléctricos $\vec{E}(\vec{x}, t)$ y magnéticos $\vec{B}(\vec{x}, t)$ vienen dadas, en unidades de Lorentz-Heaviside,[3] en la siguiente forma

$$\vec{\nabla}\cdot\vec{E} = \rho, \tag{1.21}$$

$$\vec{\nabla}\times\vec{E} = -\frac{1}{c}\,\partial_t\vec{B}, \tag{1.22}$$

$$\vec{\nabla}\cdot\vec{B} = 0, \tag{1.23}$$

$$\vec{\nabla}\times\vec{B} = \frac{1}{c}\,\vec{\jmath} + \frac{1}{c}\,\partial_t\vec{E}, \tag{1.24}$$

donde c es la velocidad de la luz.

Las ecuaciones de Maxwell en forma diferencial (1.21)-(1.24) consiste en 8 ecuaciones diferenciales lineales acopladas, donde en general se toma como condiciones de contorno que los campos $\vec{E}$ y $\vec{B}$ tienden a cero en el infinito (para sistemas infinitos). Aunque la forma diferencial (1.21)-(1.24) es más conveniente para buscar soluciones de las ecuaciones, la física detrás de estas ecuaciones se ve mejor en la formulación integral. Utilizando los teoremas de Stokes sobre las

[3] Las unidades de Lorentz-Heaviside es un convenio donde, a diferencia de las unidades SI, la permitividad eléctrica ϵ_0 y la permeabilidad magnética μ_0 del vacío no aparecen explícitamente. Tiene la ventaja que el único parámetro físico que aparece en las ecuaciones de Maxwell es la velocidad de la luz c.

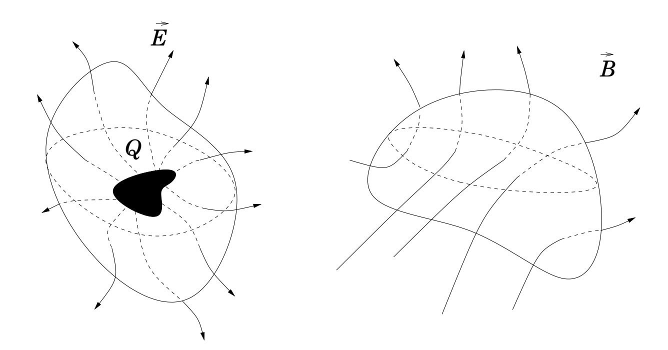


Figura 1.2: *La primera y la tercera ley de Maxwell: el flujo del campo eléctrico $\vec{E}$ a través de una superficie cerrada es proporcional a la carga total encerrada dentro de la superficie, mientras que el flujo del campo magnético $\vec{B}$ a través de cualquier superficie cerrada es cero.*

integrales de la divergencia y el rotacional de un campo vectorial $\vec{A}$,

$$\iiint_V \vec{\nabla} \cdot \vec{A} \, d^3x \;=\; \oiint_{\partial V} \vec{A} \cdot \vec{n} \, d^2x, \qquad \iint_S (\vec{\nabla} \times \vec{A}) \cdot \vec{n} \, d^2x \;=\; \oint_{\partial S} \vec{A} \cdot d\vec{r}, \tag{1.25}$$

podemos reescribir las ecuaciones de Maxwell (1.21)-(1.24) en su forma integral:

$$\oiint_{\partial V} \vec{E} \cdot \vec{n} \, d^2x \;=\; \iiint_V \rho \, d^3x, \tag{1.26}$$

$$\oint_{\partial S} \vec{E} \cdot d\vec{r} \;=\; -\frac{1}{c}\frac{d}{dt} \iint_S \vec{B} \cdot \vec{n} \, d^2x, \tag{1.27}$$

$$\oiint_{\partial V} \vec{B} \cdot \vec{n} \, d^2x \;=\; 0, \tag{1.28}$$

$$\oint_{\partial S} \vec{B} \cdot d\vec{r} \;=\; \frac{1}{c} \iint_S \vec{j} \cdot \vec{n} \, d^2x \;+\; \frac{1}{c}\frac{d}{dt} \iint_S \vec{E} \cdot \vec{n} \, d^2x. \tag{1.29}$$

La primera ley de Maxwell o la *Ley de Gauss* (1.26) nos dice que el flujo del campo eléctrico a través de una superficie cerrada es igual a la carga total encerrada dentro de la superficie. No es nada más que una versión del teorema de Gauss sobre flujos de campos vectoriales, aplicada a $\vec{E}$, identificando la *carga eléctrica*

$$Q = \int \rho \, d^3x \tag{1.30}$$

como fuente del campo eléctrico. En particular, si tomamos el caso de una carga puntual q en el origen, podemos calcular el flujo de $\vec{E}$ a través de una esfera de radio R_0 alrededor del origen. Aprovechándonos de la simetría esférica del problema es fácil de ver que (ejerc.)

$$\vec{E} \;=\; \frac{q}{4\pi r^2} \, \vec{e}_r, \tag{1.31}$$

con lo que hemos recuperado la ley de Coulomb .

La tercera ley de Maxwell (1.28) nos intenta dar el equivalente magnético de la ley de Gauss. Sin embargo aquí vemos que el lado derecho de la ecuación, que correspondería al término de la fuente del campo magnético, es cero. La ecuación (1.28) dice que el flujo magnético a través de cualquier superficie cerrada siempre es cero. En otras palabras, no existe un equivalente magnético de una carga eléctrica.Mientras que las cargas eléctricas positivas y negativas pueden existir separadas e independientes unas de otras, los polos norte y sur de un imán siempre aparecen en parejas. En otras palabras, no existen pozos y fuentes para el campos magnéticos y las líneas de

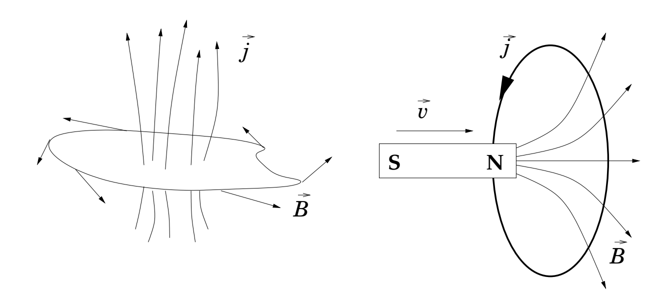


Figura 1.3: *La Ley de Ampère y la Ley de Faraday: Una corriente eléctrica a través de una superficie genera un campo magnético a lo largo de la curva que bordea la superficie, mientras que la variación de flujo magnético a través de una superficie induce una corriente eléctrica a lo largo de la curva.*

campo o bien son cerradas, o se extienden hasta el infinito. Aunque en la física moderna existen soluciones de monopolos magnéticos con propiedades muy interesantes, no hay (de momento) ninguna indicación de que estas soluciones corresponden a objetos reales en la naturaleza. Discutiremos un ejemplo de una solución de monopolo en la sección 1.6.

La *Ley de Ampère*, el primer término de la derecha de la ecuación (1.29), nos cuenta cuál es la fuente del campo magnético: la corriente eléctrica $\vec{\jmath}$. La integral de $\vec{B}$ a lo largo de una curva cerrada C es igual al flujo de corriente eléctrica a través de la superficie cuyo contorno es la curva C (Véase Figura 1.3a). En particular podemos distinguir dos casos de interés. El primero es el caso del conductor lineal infinito con una corriente constante $I = \int \vec{\jmath} \cdot \vec{n}\ d^2x$ (Figura 1.4a). A través de la ley de Ampèrey la simetría cilíndrica del sistema no es muy difícil de ver (ejerc.) que el campo magnético generado por esta corriente I tiene sólo una componente en la dirección φ, el ángulo polar en coordenadas cilíndricas:

$$\vec{B} = \frac{I}{2\pi c r}\, \vec{e}_\varphi. \tag{1.32}$$

El segundo caso interesante es el del solenoide infinito, donde una corriente I pasa por una hélice conductor (Figura 1.4b). De manera similar al caso del conductor lineal se puede demostrar (ejerc.) que el campo magnético fuera del solenoide es cero, mientras que dentro hay un campo constante $\vec{B} = I\vec{e}_z$ a lo largo del eje del solenoide. Una consecuencia de esto es que un metal se convierte en un imán si lo metemos dentro de un solenoide. Dependiendo del tipo de material es posible que las propiedades magnéticas perduren incluso si sacamos el metal del solenoide. Los casos aquí comentados son ejemplos sencillos de la *ley de Biot-Savart* sobre el campo magnético generada por corrientes eléctricos. Vemos por lo tanto que la ley de Biot-Savart, y en general la de Ampère, relacionan claramente los fenómenos eléctricos con los fenómenos magnéticos.

La *Ley de Faraday* o la *ley de la inducción magnética* (1.27) describe en cierto sentido la situación contraria a la ley de Ampère con el solenoide. Si una corriente circular causa un campo magnético, ¿también un imán metido en un anillo conductor causa una corriente eléctrica en el anillo? Esto claramente no es verdad, puesto que violaría de manera brutal la conservación de energía. Pero Faraday descubrió que sí se genera una corriente en el momento en que el imán se acerca o se aleja del anillo, es decir cuando el flujo magnético a través del anillo cambia. La ley de Faraday dice que un cambio en el flujo magnético a través de una superficie S induce un campo eléctrico rotacional alrededor de la curva que bordea S (Véase Figura 1.3b). Este campo eléctrico rotacional general una corriente $\vec{\jmath}$ que a su vez causa a través de la ley de Ampère un campo magnético. El sentido de la corriente $\vec{\jmath}$ es tal que el campo magnético causado intenta compensar el cambio de flujo del campo original. Las corrientes eléctricas inducidas por cambio de flujos magnéticos a través de una superficie son el principio básico detrás de una dinamo y forman la base del motor eléctrico.

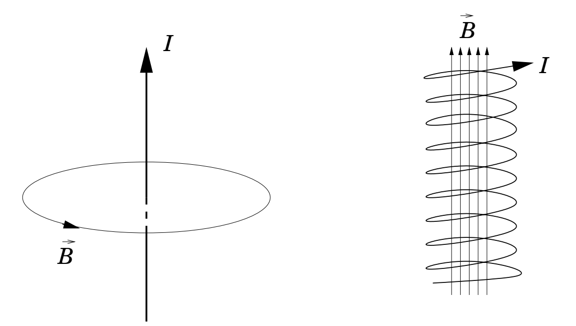


Figura 1.4: *Ejemplos típicos de la Ley de Ampère: un campo magnético rotacional alrededor de un conductor lineal infinito y un campo magnético constante en el interior de un solenoide.*

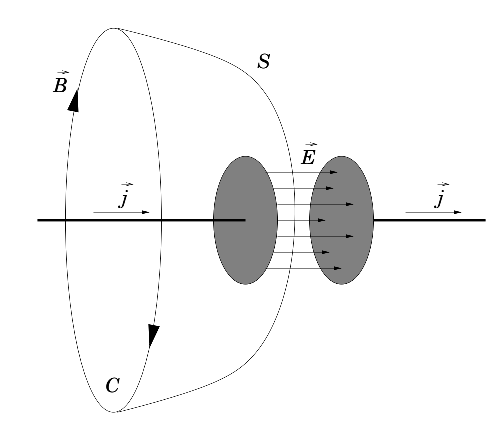


Figura 1.5: *El término de Maxwell como corrección a la ley de Ampère: el campo magnético $\vec{B}$ a lo largo de de una curva C alrededor del conductor con una corriente $\vec{\jmath}$ no es igual al flujo de la corriente pasando por la superficie S, si S pasa entre las dos placas de un condensador. La carga que se acumula en el condensador crea un cambio en el campo eléctrico, que es el responsable del campo magnético $\vec{B}$.*

Las cuatro leyes de electricidad y magnétismo que acabamos de comentar arriba, son las que eran conocidas en los tiempos de Maxwell. En ese momento eran leyes empíricas que describían bien los experimentos realizados hasta entonces. Sin embargo Maxwell se dio cuenta de que en particular la ley de Ampère (las ecuaciones (1.24) y (1.29) sin el segundo término de la derecha) no podía ser válida siempre. Si aplicamos la ley de Biot-Savart (1.32) al caso donde una corriente $\vec{\jmath}$ está cargando un condensador, la corriente genera claramente un campo magnético $\vec{B}$ alrededor del conductor. Sin embargo, el flujo de corriente a través de una superficie es cero, si dejamos que la superficie pase entre las dos placas del condensador (Véase figura 1.5), por lo que la ley de Ampère no es válida. Maxwell se dió cuenta que mientras corría la corriente eléctrica, iba aumentando la carga acumulada en el condensador y por lo tanto aumentando el flujo de campo eléctrico entre las dos placas y la superficie. De allí dedujo Maxwell que ese cambio de flujo eléctrico inducía el campo magnético $\vec{B}$ alrededor del conductor, lo que resulta en el segundo término de (1.24) y (1.29), el *término de Maxwell*. Al conjunto de la ley de Ampère y el término de Maxwell, se le suele llamar la *ley de Ampère-Maxwell*.

Maxwell introdujo su término a base de razones puramente teóricas, ya que en este momento no había ninguna indicación experimental para creer en la existencia de una contribución de este tipo. Sin embargo el término resulta ser indispensable para la conservación de la carga eléctrica y da lugar a la existencia de las soluciones de ondas electromagnéticas, un hecho experimental-

mente encontrado por Heinrich Hertz (1847-1894) alrededor de 1890.

Las leyes de Maxwell describen cómo cargas y corrientes eléctricas generan campos eléctricos y magnéticos, pero no hemos dicho todavía cómo una partícula cargada interacciona con estos campos electromagnéticos. Para esto recurrimos a la *fuerza de Lorentz*. La fuerza ejercida por un campo $\vec{E}$ y un campo $\vec{B}$ sobre una partícula con carga q viene dada por

$$\vec{F} = q\Big(\vec{E} + \frac{\vec{v} \times \vec{B}}{c}\Big), \tag{1.33}$$

donde $\vec{v}$ es la velocidad de la partícula. Vemos que esta fuerza consiste en dos partes: un término eléctrico que básicamente resume la ley de Coulomb entre dos cargas y un término que dice que la carga notará el campo magnético si se mueve con una velocidad $\vec{v}$. En particular en un campo magnético constante, una carga q con velocidad $\vec{v}$ perdendicular a $\vec{B}$ se moverá en círculos de radio $r = mv/qB$ (ejerc.). La fuerza de Lorentz es conservativa en el caso de cargas estáticas, pero no en general, ya que se ve fácilmente de las leyes de Maxwell que $\vec{\nabla} \times F \neq 0$. Sin embargo, el segundo término no realiza ningún trabajo sobre la partícula. Aunque cambia la dirección de $\vec{v}$, no cambia la norma y por lo tanto la energía cinética sigue igual. Efectivamente, el cambio de energía cinética $E_{\rm cin}$ de una partícula en un campo electromagnético sólo es debido al campo eléctrico:

$$\frac{dE_{\rm cin}}{dt} = \vec{v} \cdot \vec{F} = q\vec{v} \cdot \vec{E}. \tag{1.34}$$

1.3. Cantidades conservadas

Las mismas leyes de Maxwell indican por su propia estructura que existen ciertas cantidades conservadas. Tomando la divergencia de la Ley de Ampère-Maxwell (1.24) tenemos que

$$0 = c\vec{\nabla} \cdot (\vec{\nabla} \times \vec{B}) = \vec{\nabla} \cdot \vec{\jmath} + \vec{\nabla} \cdot \partial_t \vec{E} = \vec{\nabla} \cdot \vec{\jmath} + \partial_t \rho, \tag{1.35}$$

donde en la última igualdad hemos utilizado la ley de Gauss (1.21). La expresión (1.35) es la forma en que en una teoría de campos se expresa una ley de conservación, en particular aquí tenemos la *ley de conservación de carga eléctrica*. En la formulación integral se lee

$$\oiint_{\partial V} \vec{\jmath}\, d^2x = -\frac{d}{dt} \iiint_V \rho\, d^3x. \tag{1.36}$$

En otras palabras, el cambio de carga en un volumen V es igual al flujo de corriente que pasa por la superficie que bordea el volumen V. Si tomamos para el volumen V el universo entero, vemos que la carga total en el universo $Q = \int \rho\, d^3x$ está conservada, ya que por definición no hay flujo de corriente eléctrica a través de "la frontera de universo". Pero la ley de conservación en su forma (1.35) ó (1.36) es más fuerte que una tradicional ley de conservación global, por su carácter local. No sólo nos dice que la carga eléctrica está conservada, sino también nos dice cómo está conservada, en el sentido que indica a dónde ha ido la parte que falta: la posible pérdida de carga es debido al flujo de corriente eléctrica a través de la frontera. Nótese que para poder derivar esta ley de conservación de carga hemos utilizado explícitamente el término de Maxwell en (1.24).

Por (1.34) sabemos que el campo eléctrico puede cambiar la energía cinética de una partícula cargada. Si creemos en la ley de conservación de la energía, debemos por lo tanto aceptar que el campo electromagnético también tiene energía, que pierde o gana al aumentar o disminuir la energía de la partícula. Esto tampoco resulta tan raro, puesto que si consideramos el campo electromagnético como una realidad física, no es de extrañar que posea propiedades físicas como energía o momento.

Podemos derivar la expresión para la energía del campo electromagnético de las propias leyes de Maxwell. Si tomamos el producto escalar de (1.24) con $\vec{E}$ y de (1.22) con $\vec{B}$ tenemos

$$\begin{aligned}\vec{E}\cdot(\vec{\nabla}\times\vec{B}) &= \frac{1}{c}\vec{\jmath}\cdot\vec{E} + \frac{1}{c}\vec{E}\cdot\partial_t\vec{E},\\ \vec{B}\cdot(\vec{\nabla}\times\vec{E}) &= -\frac{1}{c}\vec{B}\cdot\partial_t\vec{B}\end{aligned} \tag{1.37}$$

Restando estas dos expresiones y utilizando que $\vec{\nabla}\cdot(\vec{A}\times\vec{B}) = \vec{B}\cdot\vec{\nabla}\times\vec{A} - \vec{A}\cdot\vec{\nabla}\times\vec{B}$ para cualesquiera dos vectores $\vec{A}$ y $\vec{B}$, vemos que

$$\begin{aligned}c\,\vec{\nabla}\cdot(\vec{E}\times\vec{B}) &= -\frac{1}{2}\,\partial_t(E^2+B^2) - \vec{\jmath}\cdot\vec{E}\\ &= -\frac{1}{2}\,\partial_t(E^2+B^2) - \sum_a q_a\vec{v}_a\cdot\vec{E}\\ &= -\frac{1}{2}\,\partial_t\Big(E^2+B^2\Big) - \frac{d}{dt}E_{\rm cin},\end{aligned} \tag{1.38}$$

donde en la segunda igualdad hemos escrito la corriente $\vec{\jmath}$ como el conjunto de cargas en movimiento y en la última igualdad hemos utilizado (1.34). La expresión tiene la forma de una ley de conservación: si identificamos las cantidades

$$E_{\rm em} = \frac{1}{2}(E^2+B^2), \qquad\qquad \vec{S} = c\,(\vec{E}\times\vec{B}), \tag{1.39}$$

con la energía del campo electromagnético y con el flujo de energía respectivamente, vemos que la ley de conservación de energía dice que el cambio de energía cinética de las partículas cargadas más el cambio de energía de campo electromagnético en un volumen es igual al flujo de energía a través de la superficie:

$$\frac{d}{dt}E_{\rm cin} + \partial_t E_{\rm em} + \vec{\nabla}\cdot\vec{S} = 0. \tag{1.40}$$

El vector $\vec{S} = c(\vec{E}\times\vec{B})$ se llama el *vector de Poynting*, llamado por John H. Poynting (1852 - 1914), un estudiante de Maxwell, que derivó esta ley de conservación de energía en 1884.

Para obtener una expresión para el momento podemos hacer una derivación parecida. Lo más cómodo es hacer la derivación en componentes. Tomando en cuenta que en componentes el producto vectorial de dos vectores $\vec{A}$ y $\vec{B}$ se escribe como

$$(\vec{A}\times\vec{B})_i = \sum_{j,k=1}^{3}\epsilon_{ijk}A_jB_k, \tag{1.41}$$

donde ϵ_{ijk} es el símbolo de Levi-Civita

$$\epsilon_{ijk} = \left\{\begin{array}{cl} 1 & \text{cuando } (ijk) \text{ es una permutación par de } (123),\\ -1 & \text{cuando } (ijk) \text{ es una permutación impar de } (123),\\ 0 & \text{en todos los demás casos,}\end{array}\right. \tag{1.42}$$

entonces vemos que la derivada parcial de la componente i del vector de Poynting con respecto

del tiempo viene dado por

$$\begin{aligned}
\frac{1}{c^2}\,\partial_t S_i &= \frac{1}{c}\sum_{j,k=1}^{3}\epsilon_{ijk}\Big[E_j\partial_t B_k \;+\; \partial_t E_j B_k\Big] \\
&= \sum_{j,k=1}^{3}\epsilon_{ijk}\Big[-E_j\sum_{l,m=1}^{3}\epsilon_{klm}\partial_l E_m \;+\; (\sum_{l,m=1}^{3}\epsilon_{jlm}\partial_l B_m - \frac{1}{c}j_j)B_k\Big] \\
&= \sum_{j,k=1}^{3}\Big[-E_j\partial_i E_j + E_j\partial_j E_i + B_k\partial_k B_i - B_k\partial_i B_k - \frac{1}{c}\epsilon_{ijk}j_j B_k\Big] \\
&= \sum_{j,k=1}^{3}\Big[-\frac{1}{2}\partial_i(E^2+B^2) + \partial_j(E_iE_j + B_iB_j) - E_i\partial_j E_j - B_i\partial_j B_j - \frac{1}{c}\epsilon_{ijk}j_j B_k\Big] \\
&= \sum_{j=1}^{3}\partial_j\Big[-\frac{1}{2}(E^2+B^2)\delta_{ij} + E_iE_j + B_iB_j\Big] - \rho E_i - \sum_{j,k=1}^{3}\frac{1}{c}\epsilon_{ijk}j_j B_k, \qquad (1.43)
\end{aligned}$$

donde en la segunda igualdad hemos utilizado las ecuaciones de Maxwell para $\vec{\nabla}\times\vec{E}$ y $\vec{\nabla}\times\vec{B}$ y en la última las ecuaciones para $\vec{\nabla}\cdot\vec{E}$ y $\vec{\nabla}\cdot\vec{B}$, aparte de la identidad

$$\sum_{i=1}^{3}\epsilon_{ijk}\epsilon_{imn} = \delta_{jm}\delta_{kn} - \delta_{km}\delta_{jn}. \qquad (1.44)$$

En los últimos dos términos de (1.43) reconocemos la expresión para la fuerza de Lorentz (1.33) de un conjunto de partículas cargadas y podemos por lo tanto escribirlo como $F_i = \frac{d}{dt}\sum_a (p_a)_i$, la derivada del momento de las partículas. Si ahora definimos el *tensor de momento electromagnético*

$$T_{ij} \;=\; (E_iE_j + B_iB_j) - \frac{1}{2}(E^2+B^2)\delta_{ij}, \qquad (1.45)$$

podemos escribir la ecuación (1.43) como

$$\partial_t\Big(c^{-2}S_i + \sum_a (p_a)_i\Big) = -\sum_{i=1}^{3}\partial_j T_{ij}. \qquad (1.46)$$

Vemos por lo tanto que podemos interpretar el vector de Poynting como el momento del campo electromagnético y la ecuación (1.46) nos dice que la pérdida de momento del campo electromagnético y la pérdida de momento de las partículas es igual al flujo del tensor de momento T_{ij}.

Nótese que el vector de Poynting aparece dos veces en estas derivaciones: una vez como el flujo de energía y otra vez como el momento conservado. En la teoría de la relatividad especial veremos que la energía y el momento están relacionados como los componentes del cuadrivector de energía-momento. Veremos en el siguiente capítulo que la teoría de Maxwell es explícitamente invariante bajo las transformaciones de Lorentz (aunque no sea visible en esta notación), por lo tanto el hecho de que el vector de Poynting aparezca en estas dos formas es una primera indicación de la covariancia de la teoría.

1.4. Potenciales electromagnéticos e invariancia gauge

Las leyes de Maxwell en su forma diferencial (1.21)-(1.24) son más cómodas de resolver que en su forma integral (1.26)-(1.29), pero aún así no son fáciles por la aparición del operador diferencial

$\vec{\nabla}$ en forma de divergencias y rotacionales. Resulta que las dos leyes de Maxwell homogéneas, la ley de Faraday (1.22) y la de la divergencia de $\vec{B}$ (1.23), nos permiten escribir el conjunto en una forma más accesible, además de darnos pistas sobre uno de las aspectos más importantes de la física moderna.

El hecho de que la divergencia de $\vec{B}$ siempre sea cero, implica, por las propiedades de divergencias y rotacionales, que podemos escribir $\vec{B}$ como el rotacional de algún campo $\vec{A}$:[4]

$$\vec{B} = \vec{\nabla} \times \vec{A}. \tag{1.47}$$

Sustituyendo este resultado el la ley de Faraday (1.22), podemos escribirla como

$$\vec{\nabla} \times \left(\vec{E} \ + \ \frac{1}{c}\partial_t \vec{A} \right) = 0, \tag{1.48}$$

intercambiando las derivadas parciales ∂_t y ∂_i. Por las mismas propiedades de rotacionales y gradientes, se ve que la combinación $\vec{E} + c^{-1}\partial_t \vec{A}$ es el gradiente de un campo $-\phi$, donde el signo menos se introduce por futura conveniencia. En otras palabras, podemos escribir el campo eléctrico como

$$\vec{E} \ = \ -\vec{\nabla}\phi \ - \ \frac{1}{c}\partial_t \vec{A}. \tag{1.49}$$

Los campos auxiliares ϕ y $\vec{A}$ son los llamados *potenciales electromagnéticos*, que nos ayudarán a escribir la teoría de Maxwell en una forma más sencilla.

Dado ϕ y $\vec{A}$, los campos electromagnéticos $\vec{E}$ y $\vec{B}$ están completamente determinados a través de las expresiones (1.47) y (1.49). Pero lo contrario no es verdad: dados $\vec{E}$ y $\vec{B}$, ni ϕ, ni $\vec{A}$ están determinados unívocamente. Si cambiamos los potenciales ϕ y $\vec{A}$ por otros potenciales ϕ' y $\vec{A}'$, relacionados con los originales como

$$\phi \rightarrow \phi' = \phi + \partial_t \Lambda, \qquad\qquad \vec{A} \rightarrow \vec{A}' = \vec{A} - c\vec{\nabla}\Lambda, \tag{1.50}$$

con $\Lambda = \Lambda(\vec{x}, t)$ una función arbitraria, entonces los campos eléctricos y magnéticos $\vec{E}'$ y $\vec{B}'$ generados por estos potenciales ϕ' y $\vec{A}'$ toman la forma

$$\begin{aligned} \vec{E}' \ &= \ -\vec{\nabla}\phi' - \frac{1}{c}\partial_t \vec{A}' \ = \ -\vec{\nabla}\Big(\phi + \partial_t\Lambda\Big) - \frac{1}{c}\partial_t\Big(\vec{A} - c\vec{\nabla}\Lambda\Big) \ = \ -\vec{\nabla}\phi - \frac{1}{c}\partial_t\vec{A} \ = \ \vec{E}, \\ \vec{B}' \ &= \ \vec{\nabla} \times \vec{A}' \ = \ \vec{\nabla} \times \Big(\vec{A} - c\vec{\nabla}\Lambda\Big) \ = \ \vec{\nabla} \times \vec{A} \ - \ c\vec{\nabla} \times \vec{\nabla}\Lambda \ = \ \vec{\nabla} \times \vec{A} \ = \ \vec{B}. \end{aligned} \tag{1.51}$$

En otras palabras, los potenciales ϕ' y $\vec{A}'$ dan exactamente los mismos campos $\vec{E}$ y $\vec{B}$ que ϕ y $\vec{A}$. Es más, dado que Λ es una función arbitraria, hay infinitos potenciales que dan todos los mismos campos electromagnéticos y por lo tanto la misma física.

¿Cómo sabemos ahora con qué ϕ y $\vec{A}$ tenemos que trabajar? ¿Cúal de estos infinitos potenciales describe correctamente la física de nuestro sistema? En realidad, la respuesta ya viene dada en (1.51): todos los potenciales ϕ y $\vec{A}$ relacionados a través de (1.50) describen la misma física, ya que todos generan los mismos campos $\vec{E}$ y $\vec{B}$. En un experimento se puede medir la intensidad y la dirección del campo electromagnético (es decir de $\vec{E}$ y $\vec{B}$), pero no existe ningún experimento que sea capaz de medir directamente ϕ y $\vec{A}$ y por lo tanto capaz de distinguir entre unos potenciales u otros relacionados a través de (1.50). En realidad, los potenciales ϕ y $\vec{A}$ son solamente unos campos auxiliares, que hemos introducido por mayor facilidad matemática, pero que no corresponden a ninguna realidad física. La realidad física está descrita únicamente por los campos $\vec{E}$ y $\vec{B}$, ya que son éstos los que medimos en los experimentos.

[4] Aquí suponemos que el espacio es simplemente conexo, lo que es el caso para $\mathbb{R}^3$. Más adelante en este curso estudiaremos soluciones en espacios que son topológicamente no-triviales.

Un ejemplo concreto ilustrará esta idea profunda: ya hemos dicho que en electrostática el campo eléctrico de una carga puntual viene dada por la ley de Coulomb (1.31),

$$\vec{E} = \frac{q}{4\pi r^2}\,\vec{e}_r, \qquad\qquad \vec{B} = 0. \tag{1.52}$$

Muchas veces se presentan los potenciales electromagnéticos de este campo eléctrico en el llamado *gauge de Coulomb* , que impone sobre el potencial vectorial $\vec{A}$ la condición $\vec{\nabla}\cdot\vec{A} = 0$. En este caso el campo eléctrico esta completamente determinado por el potencial escalar ϕ, ya que $\vec{A}$ se anula en todo el espacio:

$$\phi = -\frac{q}{4\pi r}, \qquad\qquad \vec{A} = 0. \tag{1.53}$$

Sin embargo, podemos elegir otros potenciales que nos generan el mismo campo electrostático (1.52). En particular podemos elegir el caso donde el potencial escalar es cero en todo el espacio y el campo eléctrico está soportado exclusivamente por el potencial vectorial. Efectivamente los potenciales

$$\phi' = 0, \qquad\qquad \vec{A}' = -\frac{c\,q\,t}{4\pi r^2}\,\vec{e}_r, \tag{1.54}$$

generan a través de (1.49) exactamente la misma expresión (1.52) para el campo de una carga puntual estática. A primera vista podría parecer extraño que una configuración estática como (1.52) esté descrita por un potencial que dependa explícitamente del tiempo, pero esto sólo es un prejuicio inspirado por el gauge de Coulomb, puesto que los potenciales mismos no tienen significado físico y pueden tener una dependencia temporal (lineal) sin que lo tenga $\vec{E}$. La dependencia temporal de $\vec{A}$ en (1.54) solamente es un artefacto de la elección de los potenciales, no un efecto físico. Obsérvese que efectivamente los potenciales (1.53) y (1.54) están relacionados a través de una transformación del tipo (1.50) con $\Lambda = q\,t/4\pi r$.

Por lo tanto vemos que la teoría de Maxwell tiene una cierta redundancia en su descripción: la dinámica está descrita por más variables que grados de libertad físicos y tenemos la libertad de elegir en cada momento las variables que más nos convengan. Las teorías de este tipo se llaman *teorías gauge*, la transformación (1.50) se llama una *transformación gauge* y los potenciales ϕ y $\vec{A}$ *campos gauge*.[5]

El hecho de que una transformacion gauge deje las leyes de Maxwell invariantes (ya que en éstas aparecen $\vec{E}$ y $\vec{B}$, pero no ϕ y $\vec{A}$), implica que la teoría de Maxwell tiene una simetría inesperada: dado que los potenciales no corresponden a ninguna realidad física, los verdaderos resultados físicos, como la intensidad del campo electromagnético, su energía o su momento, no pueden depender de la elección de ϕ y $\vec{A}$. Diferentes elecciones deberían dar el mismo resultado. Esto impone restricciones muy fuertes sobre la forma de las expresiones: las cantidades físicamente relevantes sólo pueden depender de las combinaciones de los potenciales ϕ y $\vec{A}$ que son invariantes gauge. En otras palabras, los potenciales sólo pueden aparecer en las cantidades físicas en las combinaciones $\vec{E} = -\nabla\phi - c^{-1}\partial_t\vec{A}$ y $\vec{B} = \vec{\nabla}\times\vec{A}$. Es fácil de comprobar que las expresiones que hemos derivado en la sección anterior satisfacen este criterio.

Es difícil sobreestimar la importancia de la invariancia gauge en la física moderna: la transformación (1.50) en la teoría de Maxwell sólo es el caso más sencillo de invariancia gauge: en 1954 los físicos (chino-)americanos Chen Ning Yang (1922 -) y Robert Mills (1927 - 1999) generalizaron la transformación (1.50) a grupos no-Abelianos y hoy en día entendemos la interacción electrodébil y la interacción fuerte (e incluso hasta cierto punto la gravedad) como teorías gauge, aunque de un tipo más complicado que en la teoría de Maxwell.

Habiendo sacado de las ecuaciones homogéneas las expresiones (1.47) y (1.49) para $\vec{E}$ y $\vec{B}$ en función de ϕ y $\vec{A}$, nos queda sustituir estas expresiones en las ecuaciones inhomogéneas restan-

[5] *To gauge* en inglés significa calibrar. La razón porque se relaciona las transformaciones (1.50) con las calibraciones es un error histórico (pero fructífero) del matemático alemán Hermann Weyl (1885 - 1955) en 1918, cuando intentó, sin éxito, unificar la relatividad general y el electromagnetismo usando invariancia bajo transformaciones de escala.

tes. Utilizando las propiedades de divergencias y rotacionales, no es difícil de ver que se puede escribir éstas como

$$-\Delta\phi \;-\; \frac{1}{c}\partial_t(\vec\nabla\cdot\vec A) \;=\; \rho,$$

$$\frac{1}{c^2}\partial_t^2\vec A \;-\; \Delta\vec A \;+\; \vec\nabla(\frac{1}{c}\partial_t\phi + \vec\nabla\cdot\vec A) \;=\; \frac{1}{c}\,\vec\jmath, \tag{1.55}$$

para la ley de Gauss (1.21) y la ley de Ampère-Maxwell (1.24) repectivamente. Aquí el operador diferencial $\Delta = \sum_{k=1}^{3}\partial_k\partial_k$ es el *laplaciano*. A primera vista estas relaciones no parecen mucho más simples que las ecuaciones de Maxwell de donde las hemos derivado, sobre todo debido a los últimos términos en el lado izquierdo. Si no fuera por este término, la segunda ecuación de (1.55) se convertiría en una ecuación de ondas inhomogénea.

Sin embargo la invariancia gauge nos proporciona la libertad de poder cambiar los potenciales ϕ y $\vec A$ según (1.50), de modo que podemos elegir los potenciales de forma que satisfagan la condición

$$\vec\nabla\cdot\vec A \;+\; \frac{1}{c}\partial_t\phi \;=\; 0, \tag{1.56}$$

tal que las ecuaciones (1.55) se reducen a

$$\frac{1}{c^2}\partial_t^2\phi \;-\; \Delta\phi \;=\; \rho, \qquad\qquad \frac{1}{c^2}\partial_t^2\vec A \;-\; \Delta\vec A \;=\; \frac{1}{c}\,\vec\jmath, \tag{1.57}$$

o sea la ecuación de ondas inhomogénea en tres dimensiones.

Aquí hemos visto otro ejemplo más de cómo la ambigüedad de los potenciales electromagnéticos, más que ser un problema, se convierte en una gran ventaja: una buena elección de ϕ y $\vec A$ puede simplificar mucho un problema dado. Concretamente la condición (1.56) se llama el *gauge de Lorenz*. Elegir unos potenciales u otros para un problema concreto se llama *fijar el gauge* y muchas elecciones de gauge son típicas para ciertas situaciones físicas. Por ejemplo, el gauge de Coulomb, $\vec\nabla\cdot\vec A = 0$, es idóneo para problemas de electrostática, pero veremos en breve que el gauge de Lorenz es más útil en relatividad especial. Otras elecciones de gauge comunes son el *gauge temporal* $\phi = 0$, el *gauge de radiación* $\vec A = 0$ (en ausencia de cargas) y el *gauge axial* $\vec A\cdot\vec n = 0$, donde $\vec n$ es un vector de unidad en una dirección específica.

Resumiendo podemos escribir las ecuaciones de Maxwell como

$$\vec E \;=\; -\vec\nabla\phi \;-\; \frac{1}{c}\partial_t\vec A, \qquad\qquad \vec B \;=\; \vec\nabla\times\vec A, \tag{1.58}$$

$$\frac{1}{c^2}\partial_t^2\phi \;-\; \Delta\phi \;=\; \rho, \qquad\qquad \frac{1}{c^2}\partial_t^2\vec A \;-\; \Delta\vec A \;=\; \frac{1}{c}\,\vec\jmath, \tag{1.59}$$

siempre y cuando asumamos que ϕ y $\vec A$ están relacionados a través de (1.56). Las ecuaciones (1.58) nos dicen cómo los potenciales interaccionan con las cargas y las corrientes, mientras (1.59) nos da la relación entre los potenciales y los campos físicos $\vec E$ y $\vec B$. Se puede demostrar que la invariancia gauge está íntimamente ligada con la teoría matemática de formas y que las ecuaciones homogéneas (1.58) son en realidad una identidad de Bianchi, que dice que el tensor electromagnético es (localmente) exacto.

1.5. La teoría de Maxwell a través del principio variacional

De la mecánica analítica conocemos las ventajas del formalismo lagrangiano para estudiar las leyes de la física: un lagrangiano es básicamente la integral de las ecuaciones de movimiento

y permite estudiar con más facilidad las propiedades y simetrías de la teoría. En particular, el principio variacional aplicado a la teoría de Maxwell nos enseña unas cuantas cosas que no son obvias mirando sólo las leyes de Maxwell (1.21)-(1.24).

La acción de la teoría de Maxwell viene dada por

$$S_{\text{Maxwell}} \;=\; \int d^4x \,\Big[\,\frac{1}{2}E^2 \;-\; \frac{1}{2}B^2 \;-\; \rho\phi \;+\; \frac{1}{c}\vec{\jmath}\cdot\vec{A}\,\Big]. \tag{1.60}$$

A primera vista esto puede parecer una acción extraña, ya que no es cuadrática en las primeras derivadas de $\vec{E}$ y $\vec{B}$, como uno esperaría de la mecánica analítica. Es más, variando ingenuamente esta acción con respecto a los campos electromagnéticos $\vec{E}$ y $\vec{B}$, las ecuaciones de movimiento darían lugar a la condición claramente errónea $\vec{E} = \vec{B} = 0$.

La razón por qué la acción (1.60) sí es la acción correcta es que en realidad las variables dinámicas de la teoría de Maxwell no son los campos electromagnéticos $\vec{E}$ y $\vec{B}$, sino los potenciales ϕ y $\vec{A}$. A través de (1.47) y (1.49) podemos reescribir la acción (1.60) como (ejerc.)

$$\begin{aligned} S_{\text{Maxwell}} \;&=\; \int d^4x \,\Big[\,\frac{1}{2}\sum_i \partial_i\phi\partial_i\phi \;+\; \frac{1}{c}\sum_i \partial_i\phi\partial_t A_i \;+\; \frac{1}{2c^2}\sum_i \partial_t A_i\partial_t A_i \\ &\quad -\frac{1}{2}\sum_{ij}\partial_i A_j\partial_i A_j \;+\; \frac{1}{2}\sum_{ij}\partial_i A_j\partial_j A_i \;-\; \rho\phi \;+\; \frac{1}{c}\sum_i j_i A_i\,\Big] \end{aligned} \tag{1.61}$$

que sí es cuadrática en las primeras derivadas de los potenciales. Aún así el lagrangiano (1.61) en términos de los potenciales sigue teniendo unas propiedades que no estamos acostumbrados a ver en sistemas de variables discretas. Primero, que aparte del término cinético $(\partial_t A_i)^2$ para $\vec{A}$, tenemos todo tipo de términos proporcionales al gradiente, $\partial_i A_j$. Además, el potencial escalar ϕ sólo tiene términos proporcionales a su gradiente $\partial_i\phi$, pero no tiene un término cinético $(\partial_t\phi)^2$, lo que hace que en realidad ϕ no sea un campo dinámico.

En realidad la razón por que aparece un tipo de términos y no el otro es la misma: la invariancia gauge. La acción debe ser invariante bajo las transformaciones (1.50), ya que la elección de los potenciales no puede influenciar la física. Sin embargo, al escribir un término cinético $(\partial_t A_i)^2$, la invariancia gauge nos obliga a incluir también los términos proporcionales a $\partial_i A_j$ y nos prohibe un término del tipo $(\partial_t\phi)^2$. Dicho de otra manera, sólo los términos que aparecen en (1.61) combinan a la acción manifestamente invariante gauge $E^2 - B^2$ de (1.60).

La idea es que las ecuaciones de movimiento de la acción (1.61) nos proporcionan las ecuaciones de Maxwell. Efectivamente, las ecuaciones de Euler-Lagrange para ϕ y $\vec{A}$,

$$\begin{aligned} &\frac{\partial}{\partial t}\Big(\frac{\delta\mathcal{L}}{\delta(\partial_t\phi)}\Big) + \sum_i \frac{\partial}{\partial x^i}\Big(\frac{\delta\mathcal{L}}{\delta(\partial_i\phi)}\Big) - \frac{\delta\mathcal{L}}{\delta\phi} = 0, \\ &\frac{\partial}{\partial t}\Big(\frac{\delta\mathcal{L}}{\delta(\partial_t A_j)}\Big) + \sum_i \frac{\partial}{\partial x^i}\Big(\frac{\delta\mathcal{L}}{\delta(\partial_i A_j)}\Big) - \frac{\delta\mathcal{L}}{\delta A_j} = 0, \end{aligned} \tag{1.62}$$

nos proporcionan precisamente las ecuaciones (1.55) (ejerc.), que sabemos que son las ecuaciones inhomogéneas de Maxwell en términos de los potenciales. En otras palabras, la ley de Gauss (1.21) y la ley de Ampère-Maxwell (1.24) son las ecuaciones de movimiento de respectivamente ϕ y $\vec{A}$ en la acción (1.61).

Por otro lado, no se obtienen las ecuaciones homogéneas (1.22) y (1.23) de esta acción a través de un principio variacional. La gran lección que nos enseña el principio variacional en la teoría de Maxwell es que en realidad el caracter de las ecuaciones homogéneas es completamente distinto a las ecuaciones inhomogéneas. Como hemos explicado en la sección 1.4, las ecuaciones

homogéneas nos informan sobre la existencia de los potenciales ϕ y $\vec{A}$: recuérdese que (1.47) y (1.49) son las soluciones generales de éstas. En otras palabras, más que unas ecuaciones dinámicas, las ecuaciones homogéneas actúan como ligaduras para $\vec{E}$ y $\vec{B}$, permitiéndonos escribirlos en términos los potenciales ϕ y $\vec{A}$. Por lo tanto, al considerar equivalentes las acciones (1.60) y (1.61) y al reconocer ϕ y $\vec{A}$ como las verdaderas variables dinámicas de la teoría, ya hemos asumido implícitamente las ecuaciones homogéneas de Maxwell en el principio variacional. Lo curioso es que esta diferencia entre las ecuaciones homogéneas e inhomogéneas apenas es apreciable a nivel de las ecuaciones mismas, mientras que está clarísima desde el punto de vista del formalismo lagrangiano.

También es interesante mirar la acción de una partícula con masa m_0 y carga eléctrica q, interaccionando con el campo electromagnético. En particular esta acción viene dada por

$$S_{\text{partícula}} \;=\; \int dt \left[\frac{1}{2}m_0 \sum_i \dot{x}^i(t)\,\dot{x}^i(t) \;-\; q\phi(x(t)) \;+\; \frac{q}{c}\sum_i \dot{x}^i(t)A^i(x(t))\right], \tag{1.63}$$

y es fácil de ver (ejerc.) que la ecuación de Euler-Lagrange

$$\frac{d}{dt}\Big(\frac{\delta\mathcal{L}}{\delta\dot{x}^i}\Big) \;-\; \frac{\delta\mathcal{L}}{\delta x^i} \;=\; 0 \tag{1.64}$$

da lugar a

$$m_0\ddot{x}^i \;+\; q\Big[\partial_i\phi \;+\; \tfrac{1}{c}\partial_t A_i\Big] \;-\; \frac{q}{c}\sum_j \dot{x}^j\Big[\partial_i A_j - \partial_j A_i\Big] \;=\; 0. \tag{1.65}$$

Dándonos cuenta de que las expresiones entre corchetes son precisamente las combinaciones (1.49) y (1.47) (no debería sorprendernos, ¡si no la ecuación no sería invariante gauge!), podemos interpretar esta ecuación como la segunda ley de Newton para una partícula sometida a la fuerza de Lorentz (1.33),

$$m_0\,\ddot{\vec{x}} \;=\; q\Big[\vec{E} \;+\; \frac{\vec{v}\times\vec{B}}{c}\Big]. \tag{1.66}$$

Nótese que debido al acoplo de $\dot{\vec{x}}$ y $\vec{A}$ en la acción (1.63), el momento canónico conjugado de x^i no es $p^i = m_0\dot{x}^i$, sino

$$p^i \;=\; m_0\dot{x}^i \;+\; \frac{q}{c}A^i, \tag{1.67}$$

y por lo tanto el hamiltoniano viene dado por

$$H_{\text{partícula}} \;=\; \int dt\,\Big[\frac{1}{2m_0}\Big(\vec{p} - \frac{q}{c}\vec{A}(x)\Big)^2 \;+\; q\phi(x)\Big]. \tag{1.68}$$

Este hamiltoniano, con *acoplo mínimo* al campo electromagnético, se utiliza mucho en la mecánica cuántica. Al discutir el efecto Aharonov-Bohm en la sección 1.6 veremos una aplicación de esta expresión.

Finalmente, aunque hemos discutido la invariancia gauge de los términos cinéticos de (1.61), no hemos dicho nada sobre la invariancia de los términos de acoplo, ni (1.61), ni en (1.63). Puede parecer extraño que los últimos dos términos de (1.63) contengan a ϕ y $\vec{A}$ tal cuales, ya que parecen romper la invariancia gauge. Sin embargo, una transformación gauge (1.50) induce la siguiente transformación en la acción

$$\begin{aligned}\mathcal{L}_{\text{partícula}} \;\rightarrow\; \mathcal{L}'_{\text{partícula}} \;&=\; \mathcal{L}_{\text{partícula}} - q\left[\partial_t\Lambda \;-\; \frac{1}{c}\sum_i \dot{x}^i\,c\,\partial_i\Lambda\right]\\ &=\; \mathcal{L}_{\text{partícula}} - q\,\frac{d}{dt}\Lambda(t,x)\\ &\sim\; \mathcal{L}_{\text{partícula}},\end{aligned} \tag{1.69}$$

donde en la segunda igualdad hemos usado la regla de la cadena y en la tercera el hecho de que dos lagrangianos que difieren por una derivada total son físicamente equivalentes. Por lo tanto, a pesar de que los potenciales aparezcan desnudos, una transformación gauge no tiene efecto físico en (1.63), ya que sólo dan lugar a una derivada total.

Una cosa parecida, aunque no igual, ocurre con los términos de acoplo de (1.60): una transformación gauge (1.50) cambiaría la acción como

$$\begin{aligned} S_{\text{Maxwell}} \rightarrow S'_{\text{Maxwell}} &= S_{\text{Maxwell}} - \int d^4x \Big[\rho\, \partial_t \Lambda + \sum_i j_i\, \partial_i \Lambda \Big] \\ &= S_{\text{Maxwell}} + \int d^4x \Big[\partial_t \rho + \sum_i \partial_i j_i \Big] \Lambda \\ &= S_{\text{Maxwell}}, \end{aligned} \tag{1.70}$$

donde en la segunda igualdad hemos integrado por partes y en la tercera hemos quitado la segunda integral, gracias a la ley de conservación de carga (1.35). En otras palabras, la acción (1.60) es invariante gauge gracias a la conservación de carga. Pero lo contrario también es verdad: por el teorema de Noether sabemos que a cada simetría continua corresponde una cantidad conservada (y vice versa), y lo que estamos viendo aquí es que la carga eléctrica es la cantidad conservada asociada a la simetría de la acción (1.60) bajo transformaciones gauge (1.50).

1.6. Soluciones de las ecuaciones de Maxwell

Con las ecuaciones (1.57) para los potenciales no resulta difícil encontrar soluciones de las ecuaciones de Maxwell. Una gran ventaja de la estructura de las ecuaciones es que son lineales, así que cualquier combinación lineal de dos soluciones también es una solución.

En la sección 1.2 ya hemos mencionado algunas soluciones, como la de la carga puntual, el conductor lineal y el solenoide infinito. En esta sección repasaremos algunas soluciones más, que son interesantes por su relevancia física.

Ondas electromagnéticas

Una de las propiedades sorprendentes de las ecuaciones de Maxwell es que existen soluciones no-triviales en ausencia de cargas y corrientes. Las llamadas *ecuaciones del vacío* son

$$\vec{\nabla} \cdot \vec{E} = 0, \qquad \vec{\nabla} \cdot \vec{B} = 0, \tag{1.71}$$

$$\vec{\nabla} \times \vec{E} = -\frac{1}{c} \partial_t \vec{B}, \qquad \vec{\nabla} \times \vec{B} = \frac{1}{c} \partial_t \vec{E}. \tag{1.72}$$

Una primera observación es la remarcable simetría entre campos eléctricos y magnéticos en la ausencia de cargas. En la presencia de cargas esta simetría está rota debido al hecho de que existen cargas eléctricas, pero no hay monopolos magnéticos. Volveremos a este tema un poco más adelante.

Una segunda observación es que la estructura de las ecuaciones del vacío permite la existencia de paquetes de campo viajando por el espacio: en la ecuación (1.72) vemos que un cambio en el campo magnético induce un campo eléctrico por la ley de Faraday, y a su vez un cambio de campo eléctrico induce por el término de Maxwell un campo magnético. Un cambio en el campo electromagnético genera por lo tanto más cambios, resultando en paquetes de campo propagándose por el espacio independientemente de cargas y corrientes. Estos paquetes de campo se llaman *ondas electromagnéticas*.

El caso más sencillo es cuando los campos $\vec{E}$ y $\vec{B}$ sólo dependen de una coordenada, digamos x, y del tiempo. Entonces las ecuaciones (1.71)-(1.72) se reducen a

$$\partial_x E_x \;=\; \partial_x B_x \;=\; \partial_t E_x \;=\; \partial_t B_x \;=\; 0, \tag{1.73}$$

$$\partial_x E_z = \frac{1}{c}\partial_t B_y, \qquad\qquad \partial_x B_y = \frac{1}{c}\partial_t E_z, \tag{1.74}$$

$$\partial_x E_y = -\frac{1}{c}\partial_t B_z, \qquad\qquad \partial_x B_z = -\frac{1}{c}\partial_t E_y. \tag{1.75}$$

La solución de (1.73) es $E_x = C_1$ y $B_x = C_2$, con C_1 y C_2 constantes que tomaremos cero, puesto que no estamos interesados de momento en campos constantes. Las dos ecuaciones de (1.74) forman, igual que las dos de (1.75), un conjunto de ecuaciones que se desacopla de las otras. Sustituyendo la segunda ecuación en en la primera y vice versa, vemos que podemos escribir (1.74) como

$$\tfrac{1}{c^2}\partial_t^2 E_z - \partial_x^2 E_z = 0, \qquad\qquad \tfrac{1}{c^2}\partial_t^2 B_y - \partial_x^2 B_y = 0, \tag{1.76}$$

y (1.75) como

$$\tfrac{1}{c^2}\partial_t^2 E_y - \partial_x^2 E_y = 0, \qquad\qquad \tfrac{1}{c^2}\partial_t^2 B_z - \partial_x^2 B_z = 0, \tag{1.77}$$

Vemos por lo tanto que cada componente satisface una ecuación de ondas unidimensional, que tiene como solución más general

$$\begin{aligned} E_y &= f_1(x+ct) + g_1(x-ct), & E_z &= f_2(x+ct) + g_2(x-ct), \\ B_z &= c^{-1} f_1(x+ct) - c^{-1} g_1(x-ct), & B_y &= -c^{-1} f_2(x+ct) + c^{-1} g_2(x-ct), \end{aligned} \tag{1.78}$$

donde f_1, f_2, g_1 y g_2 son funciones arbitrarias de su argumento. Las funciones $f(x+ct)$ representan una onda que se propaga con velocidad c, la velocidad de la luz, en la dirección del eje x negativo, mientras las funciones $g(x-ct)$ una onda en la dirección del eje x positivo. Dado que el campo no depende de las direcciónes y ó z, el campo electromagnético toma los mismos valores en los planos $x =$ constante. Las soluciones que tienen este propiedad se llaman *ondas planas*. Nótese que la amplitud de la onda es perpendicular a la dirección de propagación. Las ondas electromagnéticas son por lo tanto ondas transversales.

Una clase importante de ondas planas son las llamadas *ondas monocromáticas*, donde las funciones f y g son senos y cosenos. La solución (1.78) es entonces de la forma

$$E_i = A_i \cos(kx \pm \omega t + \varphi_i), \qquad\qquad B_i = c^{-1} A_i \cos(kx \pm \omega t + \varphi_i), \tag{1.79}$$

donde A_i es la amplitud, k el número de onda, ω la frecuencia angular y φ_i la fase. Para que las ondas monocromáticas sean de la forma (1.78), el número de onda y la frecuencia tienen que satisfacer la *relación de dispersión* en el vacío

$$\omega = ck. \tag{1.80}$$

Desde el punto de vista físico, las ondas monocromáticas no corresponden a una realidad física, puesto que se extienden por el espacio entero y tendrían una energía infinita. Sin embargo desde el punto de vista matemático es un concepto muy útil, ya que son fáciles de manejar. Además sabemos por el análisis de Fourier que cualquier solución (1.78) se puede escribir como una superposición de ondas monocromáticas de distintas frecuencias.

La longitud de onda λ viene dada por el número de onda y la frecuencia ν está relacionada con la frecuencia angular ω mediante

$$\lambda = \frac{2\pi}{k}, \qquad\qquad \nu = \frac{\omega}{2\pi}. \tag{1.81}$$

Nótese que la conocida fórmula para la velocidad de propagación de ondas nos da efectivamente la velocidad de la luz:

$$v \;=\; \lambda\nu \;=\; \frac{\omega}{k} \;=\; c. \tag{1.82}$$

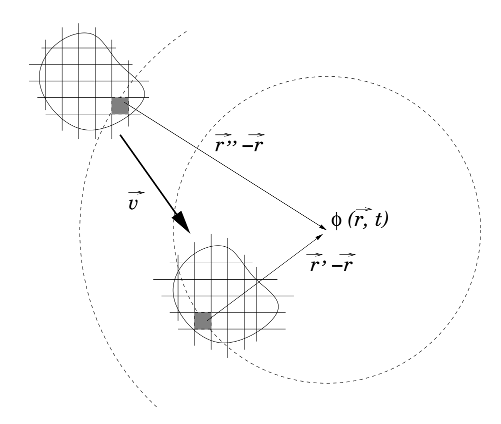


Figura 1.6: *Los potenciales retardados: los potenciales electromagnéticos en el punto $\vec{r}$ en $t = t_0$ vienen dado por la suma de todas las contribuciones de cargas y corrientes en los puntos $\vec{r}\,'$ en el momento $t - |\vec{r} - \vec{r}\,'|/c$.*

Al darse cuenta de que existen soluciones de paquetes de campo electomagnético que se propagan con la velocidad de la luz, Maxwell identificó la luz como ondas electromagnéticas. La conexión entre el electromagnetismo y la óptica fue uno de los grandes éxitos de la teoría de Maxwell. No sólo era la primera vez que alguien venía con una teoría fundamental para explicar la naturaleza de la luz, sino esa teoría también daba una base sólida a la óptica y permitía derivar las leyes empíricas de la óptica y calcular índices de relexión y refracción de materiales. Pero aparte la teoría de Maxwell es capaz de dar una explicación de cómo surgen estas ondas electromagnéticas. Esto lo veremos en el siguiente apartado.

Potenciales retardados

Resolvemos ahora las ecuaciones generales de Maxwell. Ya hemos visto que, elegiendo el gauge de Lorenz, se puede escribir las ecuaciones de Maxwell en función de los potenciales electromagnéticos como

$$\frac{1}{c^2}\partial_t^2\phi \;-\; \Delta\phi \;=\; \rho, \qquad\qquad \frac{1}{c^2}\partial_t^2\vec{A} \;-\; \Delta\vec{A} \;=\; \frac{1}{c}\,\vec{\jmath}. \tag{1.83}$$

Por la teoría de ecuaciones diferenciales sabemos que la solución general de una ecuación lineal inhomogénea consiste en la suma de las soluciones de la ecuación homogénea y una solución particular. En el caso de las ecuaciones de Maxwell, las soluciones de las ecuaciones homogéneas son las ondas electromagnéticas que hemos visto en la subsección anterior.

La derivación de la solución particular en principio no es muy diferente al caso de la electroestática o las corrientes constantes. Para calcular el valor de los potenciales en un punto $\vec{r}$ en el momento t tenemos que sumar (integrar) todas las contribuciones de cargas y corrientes en todos los elementos de volumen d^3x. La diferencia con el caso estático es que para obtener la contribución de un elemento de volumen en posición $\vec{r}\,'$ no hay que tomar la densidad de carga y corriente en el momento t, sino en el momento $t - |\vec{r} - \vec{r}\,'|/c$, dado que el tiempo $|\vec{r} - \vec{r}\,'|/c$ es el que necesita una onda electromagnética (y por lo tanto también la contribución a los potenciales) para viajar de $\vec{r}\,'$ a $\vec{r}$ (véase Figura 1.6). Los potenciales que reflejan esta propiedad se llaman los

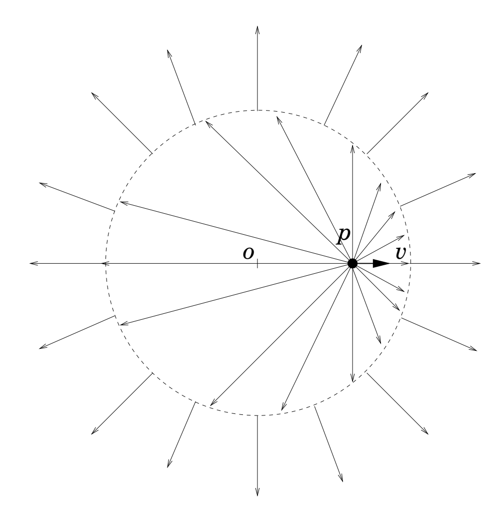


Figura 1.7: *Ondas electromagnéticas debidas a cargas aceleradas: una carga originalmente en reposo en el punto O acelera y se mueve al punto P. Lejos de la carga, las líneas de campo todavía están como si la carga estuviera en O, mientras cerca de la carga las líneas ya se han adaptado. Entremedias hay una discontinuidad que se expande con la velocidad de la luz.*

potenciales retardados y tienen la forma

$$\phi(\vec{r},t) = \frac{1}{4\pi}\iiint \frac{\rho\left(\vec{r}\,',t-\frac{|\vec{r}-\vec{r}\,'|}{c}\right)}{|\vec{r}-\vec{r}\,'|}\,d\vec{r}\,', \tag{1.84}$$

$$\vec{A}(\vec{r},t) = \frac{1}{4\pi c}\iiint \frac{\vec{\jmath}\left(\vec{r}\,',t-\frac{|\vec{r}-\vec{r}\,'|}{c}\right)}{|\vec{r}-\vec{r}\,'|}\,d\vec{r}\,'. \tag{1.85}$$

Nótese que en el caso estático recuperemos el potencial de Coulomb y el potencial de Biot-Savart.

Con la expresión de los potenciales retardados vemos lo que ya habíamos mencionado en la introducción de este capítulo: los campos electromagnéticos (o aquí los potenciales) juegan el papel de intermediarios que transmiten la interacción entre partículas. Aunque el viejo modelo newtoniano de interacciones a distancia ya parecía poco físico y causaban muchas críticas desde los primeros momentos[6], estos inconvenientes fueron olvidados rápidamente al ver el éxito de la ley de la gravedad a la hora de resolver problemas concretos. En el contexto de teoría de campos, una partícula no interacciona directamente con otra, sino con los campos presentes, que hacen propagar la influencia de la partícula con una velocidad finita (la velocidad de la luz) hasta llegar a otras partículas.

En este sentido las ondas electromagnéticas en la teoría de Maxwell son precisamente esto: distorsiones del campo electromagnético debido a la aceleración de cargas y que se propagan con la velocidad de la luz. Consideremos una carga inicialmente en reposo. En el momento $t = t_0$ la carga acelera hasta una velocidad $\vec{v}$ durante un intervalo Δt. La configuración de las líneas de campo en un momento $t > t_0 + \Delta t$ consta de dos partes (véase Figura 1.7): a distancias $r > ct$ las líneas de campo todavía están como estaban antes de que la carga se moviese, dado que la señal no ha tenido tiempo para llegar hasta allí. La contribución a los potenciales retardados en los puntos con $r > ct$ es como si la carga todavía estuviera en el sitio original. A distancias $r < c(t - \Delta t)$ sin embargo las líneas de campo apuntan directamente a la carga por la ley de Coulomb. Entre las dos zonas hay una discontinuidad esferica de grosor $c\Delta t$ que, debido a la

[6]Newton nunca pretendía tener una explicación física para la gravedad, sólo quería dar la expresión correcta de la fuerza. De allí su respuesta: "Hipotheses non fingo" (No invento hipótesis).

estructura de las ecuaciones del vacío, se expande con la velocidad de la luz, formando una onda electromagnética esférica.

La propiedad de que la información se propague con velocidad menor o igual a la de la luz es típica de teorías relativistas, dado que la velocidad de la luz como velocidad máxima es justo uno de los postulados de la teoría de la relatividad especial, como veremos en los siguientes capítulos. El hecho de que esto aparezca de manera natural en la teoría de Maxwell es otra manifestación de que esta teoría ya es relativista.

El efecto Aharanov-Bohm

Hemos mencionado en varias ocasiones que en la teoría de Maxwell los campos físicos son los campos electromagnéticos $\vec{E}$ y $\vec{B}$, mientras los potenciales ϕ y $\vec{A}$ sólo son construcciones matemáticas para manejar mejor las ecuaciones de Maxwell. Esto es obvio para la física clásica. Sin embargo, en la mecánica cuántica la cosa es un poco más sutil, ya que las partículas cargadas acoplan directamente a los potenciales. El ejemplo más conocido es el *efecto Aharonov-Bohm*, llamado así por los descubridores David Bohm (1917-1992) y su estudiante Yakir Aharonov (1932), que publicaron en 1959.[7]

En la mecánica cuántica, la norma al cuadrado $|\Psi(\vec{r},t)|^2$ de la función de onda $\Psi(\vec{r},t)$ de un electrón (o cualquier otra partícula) describe la probabilidad de encontrar el electrón en la posición $\vec{r}$ en un momento t. Nótese que por lo tanto la función de onda no está determinada unívocamente: dada una función de onda $\Psi(\vec{r},t)$, siempre se puede añadir una fase $\alpha = \alpha(\vec{r},t)$, tal que $\Psi(\vec{r},t)$ y $\Psi'(\vec{r},t) = e^{i\alpha(\vec{r},t)}\Psi(\vec{r},t)$ describen la misma física.

El hecho de que ni los potenciales electromagnéticos, ni la fase de la función de onda representen conceptos físicos, sugiere que una cosa esté relacionada con la otra. Efectivamente, a través del formalismo hamiltoniano se puede demostrar que una transformación gauge de los potenciales induce un cambio de fase en la función de onda y vice versa.

En la sección 1.5 hemos visto que el hamiltoniano clásico de una partícula con masa m y carga q en un campo electromagnético externo con potencial $\vec{A}$ viene dado por

$$H = \frac{1}{2m}(p_i - qA_i)^2 + \phi(x). \tag{1.86}$$

Entonces el hamiltoniano cuántico se obtiene remplazando los momentos p_i por $-i\hbar\partial_i$, resultando el operador

$$\hat{H} = -\frac{\hbar^2}{2m}\Big(\partial_i + \frac{iq}{\hbar}A_i\Big)^2 + \phi(x_i) \tag{1.87}$$

con $\hbar = h/2\pi$, es la constante de Planck reducida. El hamiltoniano es el operador de evolución de la función de onda $\Psi(\vec{r},t)$, ya que aparece en la ecuación de movimiento, la ecuación de Schrödinger

$$\hat{H}\Psi(\vec{r},t) = i\hbar\,\partial_t\Psi(\vec{r},t). \tag{1.88}$$

El hecho de que aparezcan explícitamente los potenciales electromagnético $\vec{A}$ y ϕ en el hamiltoniano, hace que éste no sea invariante bajo transformación gauge (1.50). Sin embargo, una transformación gauge no puede tener consecuencias físicas, por lo tanto la única manera de conseguir que la ecuación de Schrödinger sea invariante es añadirle un factor de fase $e^{icq\Lambda/\hbar}$ a la función de onda. Esta es la profunda relación entre los cambios de fase y la invariancia gauge que descubrió el físico ruso Vlamidir Fock (1898 - 1974) en 1926.

Es justo este cambio de fase el que utiliza el efecto Aharonov-Bohm. El experimento es el siguiente: considera una fuente de electrones colocada frente a una pantalla (detector). Entre la

[7] En realidad el efecto ya fue descrito en 1949 por Werner Ehrenberg y Raymond Siday, por lo que a veces se refiere a ello como el efecto Ehrenberg-Siday-Aharonov-Bohm.

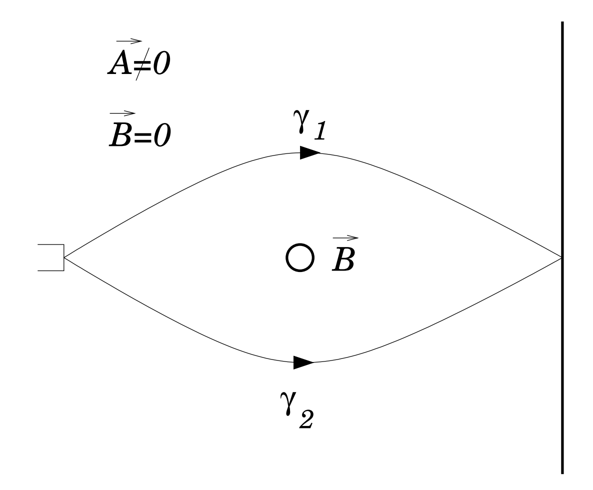


Figura 1.8: *El efecto Aharonov-Bohm: electrones emitidos por una fuente, pasan alrededor de un solenoide. El campo magnético $\vec{B}$ fuera del solenoide es cero, sin embargo el potencial $\vec{A}$ no. Clásicamente los electrones no estan influenciados por el flujo magnético dentro del solenoide. Cuánticamente la función de ondas acoge una fase, que da lugar a un patrón de interferencias en una pantalla (detector).*

fuente de electrones y la pantalla colocamos un solenoide por el cual pasa una corriente (véase figura 1.8). Ya hemos visto en la sección 1.2 que el solenoide genera un campo magnético $\vec{B} = 0$ fuera y $\vec{B} = I\vec{e}_z$ dentro del solenoide. Tomando por simplicidad el radio del solenoide a cero, manteniendo el flujo magnético $\Phi = \iint \vec{B} \cdot \vec{n}\, d^2x$ constante, el potencial que genera este campo magnético en coordenadas cilíndricas (con el eje z a lo largo del solenoide) viene dado por

$$\vec{A} = \frac{\Phi}{2\pi r}\vec{e}_\varphi. \tag{1.89}$$

Efectivamente, a través de la expresión para el rotacional en coordenadas cilíndricas,

$$\vec{\nabla} \times \vec{A} = \Big(r^{-1}\partial_\varphi A_z - \partial_z A_\varphi\Big)\vec{e}_r + \Big(\partial_z A_r - \partial_r A_z\Big)\vec{e}_\varphi + r^{-1}\Big(\partial_r(rA_\varphi) - \partial_\varphi A_r\Big)\vec{e}_z, \tag{1.90}$$

no es difícil de comprobar que con esta expresión $\vec{B}$ es cero en todo el espacio, menos en el origen $r = 0$.

Clásicamente los electrones no notan la presencia del flujo magnético, puesto que $\vec{B} = 0$ a lo largo de toda su trayectoria (y por lo tanto la fuerza de Lorentz es cero). Sin embargo en la mecánica cuántica, al conectar el flujo, la función de ondas de los electrones acoge un factor de fase

$$\Psi^{(0)}(\vec{r}) \to \Psi^{(\vec{A})} = \exp\Big(- \frac{icq}{\hbar}\int_\infty^{\vec{r}} \vec{A}(\vec{r}\,') \cdot d\vec{r}\,'\Big)\, \Psi^{(0)}(\vec{r}). \tag{1.91}$$

Debido a esta fase, el comportamiento en la pantalla será diferente. La probabilidad de encontrar el electrón en la posición $\vec{r}$ viene dado por el cuadrado de la norma de la suma de las dos contribuciones, una por cada camino γ_1 y γ_2, por encima y por debajo del solenoide respectivamente (véase figura 1.8):

$$P \sim \Big|\Psi_{\gamma_1}(\vec{r}) + \Psi_{\gamma_2}(\vec{r})\Big|^2. \tag{1.92}$$

En general la contribución de los dos caminos causará un patrón de interferencias en la pantalla, debido a que en algunos puntos las dos contribuciones se anulan, mientras en otros puntos se suman.

Ahora, aunque los electrones no notan en ningún momento el campo magnético, está claro que la probabilidad P será distinta para $\Phi = 0$ que para $\Phi \neq 0$. En general tenemos que

$$\begin{aligned} P &\sim \left| \exp\Big(-\frac{icq}{\hbar}\int_{\gamma_1} \vec{A}(\vec{r}\,')\cdot d\vec{r}\,'\Big)\, \Psi^{(0)}_{\gamma_1}(\vec{r}) \;+\; \exp\Big(-\frac{icq}{\hbar}\int_{\gamma_2} \vec{A}(\vec{r}\,')\cdot d\vec{r}\,'\Big)\, \Psi^{(0)}_{\gamma_2}(\vec{r}) \right|^2 \\ &= \left| \exp\Big(-\frac{icq}{\hbar}\int_{\gamma_2} \vec{A}(\vec{r}\,')\cdot d\vec{r}\,'\Big)\right|^2 \cdot \left| \exp\Big(-\frac{icq}{\hbar}\int_{\gamma_1-\gamma_2} \vec{A}(\vec{r}\,')\cdot d\vec{r}\,'\Big)\, \Psi^{(0)}_{\gamma_1}(\vec{r}) \;+\; \Psi^{(0)}_{\gamma_2}(\vec{r}) \right|^2 \\ &= \left| e^{-icq\Phi/\hbar}\Psi^{(0)}_{\gamma_1}(\vec{r}) \;+\; \Psi^{(0)}_{\gamma_2}(\vec{r}) \right|^2, \end{aligned} \tag{1.93}$$

donde hemos utilizado que la integral de contorno de $\vec{A}$ a lo largo de una trayectoria cerrada $\gamma = \gamma_1 - \gamma_2$ alrededor del solenoide es igual a Φ:

$$\oint_\gamma \vec{A}\cdot d\vec{r} \;=\; \iint (\vec{\nabla}\times\vec{A})\cdot\vec{n}\, d^2x \;=\; \iint \vec{B}\cdot\vec{n}\, d^2x \;=\; \Phi. \tag{1.94}$$

De (1.93) está claro que la medición en la pantalla será diferente para $\Phi = 0$ y $\Phi \neq 0$. El factor de fase extra, en el caso de que el flujo magnético sea distinto de cero, causa que el patrón de interferencias generado se desplace con respecto al caso sin flujo. Esto es francamente sorprendente, puesto que el desplazamiento del patrón de interferencias es un efecto físico, mientras que los electrones sólo han podido interaccionar con $\vec{A}$, puesto que $\vec{B} = 0$ fuera del solenoide. ¿Significa esto que el potencial $\vec{A}$ sí tiene significado físico en la mecánica cuántica? Analicemos la situación en más detalle.

Aunque el campo magnético fuera del solenoide es cero, hay claramente un efecto físico debido a la presencia de $\vec{A}$. Sin embargo, como se puede ver en (1.94), este efecto es proporcional a Φ, el flujo de $\vec{B}$ a través del solenoide, que claramente es una cantidad que es invariante gauge. Mirando bien la fórmula (1.94), hay algo raro que nos debería llamar la atención: dado que $\vec{B} = 0$, el potencial $\vec{A}$ tiene que ser un *gauge puro*,[8] es decir $\vec{A} = -c\vec{\nabla}\Lambda$, puesto que sólo así $\vec{\nabla}\times\vec{A} = 0$. Efectivamente, a partir de (1.89) y la fórmula para el gradiente en coordenadas cilíndricas

$$\vec{\nabla}\Lambda = \partial_r\Lambda\vec{e}_r \;+\; r^{-1}\partial_\varphi\Lambda\vec{e}_\varphi \;+\; \partial_z\Lambda\vec{e}_z, \tag{1.95}$$

sacamos que el potencial (1.89) es el gradiente de una función

$$\Lambda(\varphi) = -\frac{1}{2\pi c}\Phi\,\varphi, \tag{1.96}$$

en todo el espacio salvo en $r = 0$. Sin embargo, por el teorema de Stokes (1.25) la integral sobre una curva cerrada de un gradiente (o equivalentemente, el rotacional de un gradiente) es cero, mientras que la contribución en (1.94) nos da un resultado finito Φ, justo el efecto físico.

La razón por la que la integral (1.94) no es cero, a pesar de que $\vec{A}$ es un gradiente, es porque la curva γ es topológicamente no-trivial. En un espacio topológicamente trivial la integral de un gradiente sobre una curva cerrada es cero, porque se puede contraer la curva a un solo punto. Sin embargo en este caso, el potencial $\vec{A}$ diverge en $r = 0$, donde está colocado el solenoide, lo que nos impide contraer la curva γ a un punto sin que pase por la singularidad. El espacio de configuraciones en que estamos trabajando es por lo tanto $\mathbb{R}^2\backslash\{0\}$, el plano quitando el origen, y este espacio no es *simplemente conexo*: las curvas que pasan alrededor del origen no son contraíbles a un solo punto. Es justo debido a esta estructura topológica que la integral (1.94) (y por lo tanto el efecto físico Φ) no es cero.

[8] Se dice de unos potenciales ϕ y $\vec{A}$ que son puro gauge si los campos eléctricos y magnéticos generados por ellos son identicamente cero, es decir, si $\phi = \partial_t\Lambda$ y $\vec{A} = -c\vec{\nabla}\Lambda$.

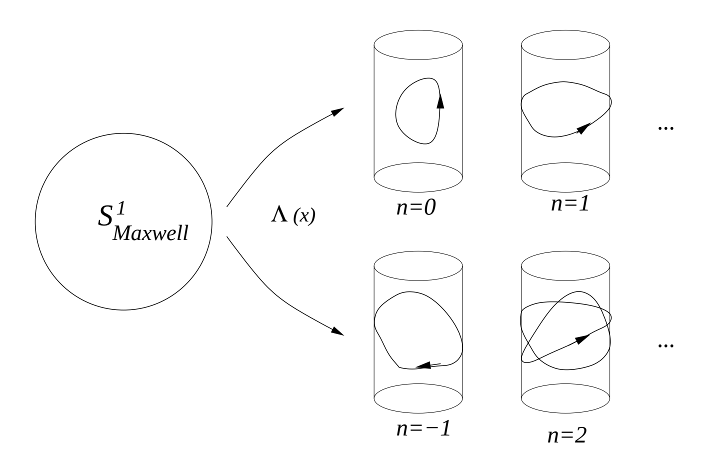


Figura 1.9: *La topología del efecto Aharonov-Bohm: el grupo gauge de la teoría de Maxwell tiene la topología de un círculo S^1 (izquierda), mientras que el espacio de configuraciones tiene la topología de un cilindro, $\mathbb{R} \times S^1$ (los 4 casos de la derecha). Una transformación gauge $\Lambda(x)$ es una función de S^1_{Maxwell} a la S^1_{config} del cilindro y las distintas transformaciones se dividen en distintas clases, caracterizadas por el número n de veces que la S^1_{Maxwell} enrolla la S^1_{config} en una dirección o en la otra. Funciones de distintas clases no son deformables unas en otras de manera continua, de modo que el número n es un invariante topológico.*

Por lo tanto, ¿cuál es al final el estatus físico del potencial en el problema de Aharonov-Bohm? ¿Corresponde a un campo "real", con efectos físicos? Como ya hemos dicho, el efecto físico es proporcional a Φ, el flujo magnético a través del solenoide, que depende de $\vec{B}$ y no de $\vec{A}$. Todavía podemos hacer las transformaciones gauge $\vec{A} \to \vec{A} - c\vec{\nabla}\Lambda$ que queramos, siempre y cuando respetemos las condiciones de contorno, es decir la estructura topológica del espacio $\mathbb{R}^2 \backslash \{0\}$. Al ser el efecto físico proporcional a $\vec{B}$ y al tener aún cierta libertad gauge, podemos decir tranquilamente que tampoco en el efecto Aharonov-Bohm los potenciales tienen significado físico.

Por último, existe una explicación matemática elegante del efecto Aharonov-Bohm en términos de estructuras topológicas. Matemáticamente hablando el efecto es debido a que ni el espacio de configuraciones, ni el grupo gauge son simplemente conexos. Por un lado el grupo gauge de la teoría de Maxwell es $U(1)$ (el grupo de multiplicación por una fase $e^{i\alpha}$), cuyo espacio de grupo es el círculo S^1. Por otro lado, el espacio de configuraciones $\mathbb{R}^2 \backslash \{0\}$, que es topológicamente igual a $\mathbb{R} \times S^1$, pero dado que (1.96) no depende de r, la parte (topológicamente) relevante del espacio de configuraciones es S^1. El parámetro Λ de una transformación gauge es por lo tanto una función del espacio de grupo S^1 al espacio (relevante) de configuraciones S^1 (véase Figura 1.9). Las diferentes funciones Λ se distinguen por el número de veces que la primera S^1 recorre (enrolla) a la segunda en una dirección o en la otra. Funciones con distinto número de enrollamiento no son continuamente deformables unas en otras y funciones con número de enrollamiento distinto de cero no son contraíbles. El número de enrollamiento (lo que físicamente corresponde al número de vueltas que da el electrón alrededor del solenoide) es por lo tanto un invariante topológico y la clasificación de las distintas funciones $\Lambda : S^1 \to S^1$ se hace a través del *(primer) grupo de homotopía* π_1. Para el caso del grupo $U(1)$, el grupo de homotopía $\pi_1(U(1)) = \pi_1(S^1) = \mathbb{Z}$.

El monopolo de Dirac

Ya hemos comentado antes que las ecuaciones del vacío (1.71)-(1.72)

$$\vec{\nabla} \cdot \vec{E} = 0, \qquad \vec{\nabla} \cdot \vec{B} = 0, \tag{1.97}$$

$$\vec{\nabla} \times \vec{E} = -\frac{1}{c}\partial_t \vec{B}, \qquad \vec{\nabla} \times \vec{B} = \frac{1}{c}\partial_t \vec{E}. \tag{1.98}$$

son remarcablemente simétricas bajo intercambio de $\vec{E}$ y $\vec{B}$. Efectivamente en ausencia de cargas las ecuaciones (1.98) son invariantes bajo la transformación $SO(2)$

$$\vec{E}\,' = \cos\alpha\,\vec{E} + \sin\alpha\,\vec{B}, \tag{1.99}$$
$$\vec{B}\,' = -\sin\alpha\,\vec{E} + \cos\alpha\,\vec{B}. \tag{1.100}$$

Nótese que esta transformación intercambia completamente $\vec{E}$ y $\vec{B}$, de modo que en el vacío los campos eléctricos y magnéticos son en realidad indistinguibles y que es puramente convenio lo que llamamos $\vec{E}$ y lo que llamamos $\vec{B}$.

Esta simetría se rompe en presencia de cargas, debido al hecho de que existen cargas eléctricas, pero no hay cargas magnéticas. El físico inglés Paul Dirac (1902-1984) descubrió en 1931 una configuración de campos que describe un monopolo magnético.

Para restaurar la simetría en presencia de cargas y corrientes eléctricas ρ_e y $\vec{\jmath}_e$, es preciso introducir cargas magnéticas ρ_m y corrientes magnéticas $\vec{\jmath}_m$ en las leyes de Maxwell:

$$\vec{\nabla}\cdot\vec{E} = \rho_e, \qquad \vec{\nabla}\cdot\vec{B} = \rho_m,$$
$$\vec{\nabla}\times\vec{E} = -\frac{1}{c}\,\vec{\jmath}_m - \frac{1}{c}\partial_t\vec{B}, \qquad \vec{\nabla}\times\vec{B} = \frac{1}{c}\,\vec{\jmath}_e + \frac{1}{c}\partial_t\vec{E}. \tag{1.101}$$

En analogía con la expresión (1.31) de una carga puntual, un monopolo magnético con carga magnética q_m en el origen produciría un campo magnético $\vec{B}$ de la forma

$$\vec{B} = \frac{q_m}{4\pi r^2}\vec{e}_r, \tag{1.102}$$

Un potencial $\vec{A}$ que daría lugar al campo (1.102) viene dado por

$$\vec{A}^{(n)} = \frac{q_m(1-\cos\theta)}{4\pi r\sin\theta}\,\vec{e}_\varphi, \tag{1.103}$$

ya que con la expresión

$$\vec{\nabla}\times\vec{A} = \frac{1}{r\sin\theta}\Big(\partial_\theta(\sin\theta A_\varphi) - \partial_\varphi A_\theta\Big)\vec{e}_r + \Big(\tfrac{1}{r\sin\theta}\partial_\varphi A_r - \tfrac{1}{r}\partial_r(rA_\varphi)\Big)\vec{e}_\theta$$
$$+\frac{1}{r}\Big(\partial_r(rA_\theta) - \partial_\theta A_r\Big)\vec{e}_\varphi, \tag{1.104}$$

para el rotacional en coordenadas esféricas, se puede comprobar que $\vec{\nabla}\times\vec{A}^{(n)} = \vec{B}$ en todo el espacio, salvo en $r = 0$. Nótese que la expresión (1.103) es singular para $\theta = \pi$, es decir el eje z negativo. Esta singularidad, llamada la *cuerda de Dirac*, es una consecuencia del hecho de que hayamos modificado la expresión para la divergencia de $\vec{B}$: sólo a través de una singularidad en $\vec{A}$ podemos escribir

$$\vec{\nabla}\cdot\vec{B} = \vec{\nabla}\cdot(\vec{\nabla}\times\vec{A}) \neq 0. \tag{1.105}$$

Nótese que esta situación es muy parecida a la de (1.94) en el caso del efecto Aharonov-Bohm, donde también los teoremas del análisis vectorial son circunvalados por singularidades en el potencial. Esto sugiere que también aquí habrá un fundamento topológico en la base de esta solución. Veremos que éste efectivamente es el caso.

A nivel clásico la singularidad de la cuerda de Dirac no nos debe preocupar, puesto que sólo es una singularidad en el potencial y no corresponde a nada físico. Sin embargo ya hemos visto que a nivel cuántico sí puede haber efectos físicos. Por lo tanto si queremos tomar el monopolo de Dirac en serio, tenemos que encontrar una manera de deshacernos de la cuerda de Dirac.

Podemos evitar la singularidad de la cuerda de Dirac en $\theta = \pi$, usando otra expresión para el potencial $\vec{A}$. Una expresión para $\vec{A}$ que da el mismo campo magnético (1.102) es

$$\vec{A}^{(s)} = \frac{-q_m(1+\cos\theta)}{4\pi r\sin\theta}\,\vec{e}_\varphi, \tag{1.106}$$

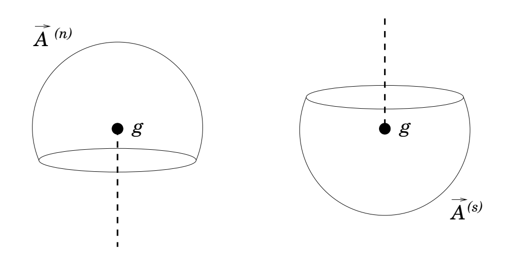


Figura 1.10: *El monopolo de Dirac según Wu y Yang: dividimos el espacio en dos parches* $\theta \in [0, \frac{\pi}{2} + \epsilon]$ *y* $\theta \in [\frac{\pi}{2} - \epsilon, \pi]$ *y utilizamos en cada parche los potenciales* $\vec{A}^{(n)}$ *y* $\vec{A}^{(s)}$ *respectivamente, que son regulares en su respectivos parches. La cuerda de Dirac (la línea interrumpida) no aparece en la descripción, ya que es un artefacto de la elección de gauge.*

que efectivamente evita la singularidad en $\theta = \pi$, pero que diverge en $\theta = 0$, el eje z positivo. El hecho de que la singularidad cambie de sitio al elegir otro potencial, indica que la cuerda de Dirac no es algo físico, sino un artefacto de la elección de gauge.[9] Efectivamente, $\vec{A}^{(n)}$ y $\vec{A}^{(s)}$ están relacionados a través de una transformación gauge (puesto que dan el mismo $\vec{B}$):

$$\vec{A}^{(n)} \;=\; \vec{A}^{(s)} + \frac{q_m}{2\pi r \sin\theta}\vec{e}_\varphi \;=\; \vec{A}^{(s)} - c\vec{\nabla}\Big(-\frac{q_m}{2\pi c}\,\varphi\Big), \tag{1.107}$$

donde en la última igualdad hemos utilizado la expresión

$$\vec{\nabla}\Lambda \;=\; \partial_r\Lambda\,\vec{e}_r \;+\; \frac{1}{r}\,\partial_\theta\Lambda\,\vec{e}_\theta \;+\; \frac{1}{r\sin\theta}\,\partial_\varphi\Lambda\,\vec{e}_\varphi, \tag{1.108}$$

para el gradiente en coordenadas esféricas.

El truco desarrollado por los físicos chinos Chien-Shiung Wu (1912-1997) y Chen Ning Yang (el mismo de la invariancia gauge tipo Yang-Mills) consiste en dividir el espacio en parches y utilizar un potencial en cada parche, tal que la singularidad del potencial cae fuera del parche en que lo utilizas (véase figura 1.10). En particular cogemos como primer parche el hemisferio norte $\theta \in [0, \frac{\pi}{2} + \epsilon]$, donde utilizamos el potencial $\vec{A}^{(n)}$ y como segundo parche el hemisferio sur $\theta \in [\frac{\pi}{2} - \epsilon, \pi]$ con el potencial $\vec{A}^{(s)}$. En la zona de solapamiento $\theta \in [\frac{\pi}{2} - \epsilon, \frac{\pi}{2} + \epsilon]$ los dos potenciales están relacionados a través de la transformación gauge (1.107). Nótese que esta transformación no está bien definida en $\theta = 0$ y $\theta = \pi$, los puntos de la cuerda de Dirac, pero dado que sólo estamos interesados en aplicar la transformación gauge en la zona de solapamiento, no nos tiene que preocupar.

El flujo magnético total a través de una esfera alrededor del monopolo viene dado por

$$\begin{aligned}
\oint\!\!\oint \vec{B}\cdot\vec{n}d^2x \;&=\; \iint_N (\vec{\nabla}\times\vec{A}^{(n)})\cdot\vec{n}\,d^2x \;+\; \iint_S (\vec{\nabla}\times\vec{A}^{(s)})\cdot\vec{n}\,d^2x \\
&=\; \oint \Big(\vec{A}^{(n)} \;-\; \vec{A}^{(s)}\Big)\cdot d\vec{r} \\
&=\; \oint \vec{\nabla}\Big(\frac{q_m\,\varphi}{2\pi}\Big)\cdot d\vec{r} \;=\; q_m,
\end{aligned} \tag{1.109}$$

donde en la primera y la segunda igualdad hemos utilizado los teoremas de Stokes (1.25). Las integrales de superficie están tomadas respectivamente en el hemisferio norte y sur y la integral de contorno en la segunda igualdad está tomado a lo largo del ecuador, donde el signo relativo es debido al hecho de que recorremos la curva dos veces en sentidos opuestos. Finalmente en la

[9] Incluso se puede demostrar relativamente fácil con la ley de Gauss que la cuerda no tiene porqué ser recta. Puede tener la forma de una curva semi-infinita, mientras que sea continua.

última igualdad hemos utilizado las expresiones (1.107) y (1.108) para el gradiente. La ecuación (1.109) es por lo tanto la ley de Gauss para cargas magnéticas: el flujo total a través de una superficie cerrada es igual a la carga encerrada dentro de la superficie, en acuerdo con la solución (1.102) propuesta al principio.

Con esta construcción la singularidad de la cuerda de Dirac no aparece en la descripción, demostrando que no es más que un artefacto de la elección de gauge. Clásicamente el procedimiento de utilizar potenciales distintos en regiones distintas del espacio es directo y sin consecuencias. Cuánticamente es más sutil, puesto que antes hemos visto que al cambiar de gauge, la función de onda de un electrón adquiere una fase. Por lo tanto, las funciones de onda que describen el electrón con carga q_e en el hemisferio norte y sur están relacionadas mediante

$$\Psi^{(n)}(\vec{r}) = \exp\Big(-\frac{iq_eq_m}{2\pi\hbar}\,\varphi\Big)\Psi^{(s)}(\vec{r}). \tag{1.110}$$

Sin embargo, este cambio de fase tiene consecuencias profundas: obviamente en el ecuador $\theta = \pi/2$, la función de onda tiene que ser univaluada al dar una vuelta alrededor de la esfera. En otras palabras, la fase en (1.110) ha de tener el mismo valor para $\varphi = 0$ y $\varphi = 2\pi$, lo que implica que su argumento $q_eq_m/2\pi\hbar$ tiene que ser un número entero n. Escrito de otra forma, llegamos a la llamada *condición de cuantización de carga* de Dirac:[10]

$$q_e = \frac{2\pi n\hbar}{q_m}. \tag{1.111}$$

En otras palabras, una consecuencia directa de la existencia de monopolos magnéticos es que las cargas eléctricas aparecen cuantizadas, en multiplos enteros de una carga mínima $q_e = 2\pi\hbar/q_m$. Ni siquiera es necesario que el monopolo magnético esté cerca, con que hubiera un solo monopolo magnético en alguna parte del universo, todas las cargas eléctricas estarían cuantizadas. Esto es un resultado francamente sorprendente, ya que a pesar de que nunca se haya visto (con seguridad) un monopolo magnético, la cuantización de carga es precisamente lo que ocurre en la naturaleza: todas las partículas conocidas tienen una carga que es un multiplo entero de la carga del electrón.[11] Por ejemplo, aunque el electrón y el protón son dos partículas muy distintas, la cota experimental de la diferencia de sus cargas es

$$\frac{|q_{e^-} + q_p|}{e} < 1,0\cdot 10^{-21}. \tag{1.112}$$

Obviamente esto no puede servir como prueba de la existencia de monopolos, pero por otro lado no se conoce ningún otro mecanismo convincente que pueda explicar la cuantización de la carga. Es más, en 1974 el físico holandés Gerard 't Hooft (1946) y el ruso Alexander Polyakov (1945) intentaron cuantizar la carga eléctrica por otros medios, embebiendo el electromagnetismo en una teoría gauge mas grande, pero encontraron que los llamados *monopolos de 't Hooft-Polyakov* aparecen por la puerta trasera como defectos topológicos, al romper la simetría para recuperar la teoría de Maxwell. En otras palabras, parece que (de momento) no hay manera de cuantizar la carga eléctrica sin de algún modo introducir monopolos magnéticos también.

Igual que en el caso del efecto Aharonov-Bohm, la base matemática del monopolo de Dirac es una estructura topológica no-trivial. El espacio de configuraciones es $\mathbb{R}^3\backslash\{0\}$, puesto que el potencial es divergente en $\vec{r} = 0$. (Hemos eliminado el resto de la cuerda de Dirac con el truco de Wu y Yang, pero la singularidad en el origen aparece en ambos potenciales). Estamos por lo tanto trabajando con un fibrado $U(1)$ sobre el espacio $\mathbb{R}^3\backslash\{0\}$, que tiene la misma homotopía que S^2. La S^2

[10]La misma formula se puede derivar también utilizando un sólo potencial en el espacio entero. Para evitar un efecto Aharonov-Bohm visible alrededor de la cuerda de Dirac, hay que ajustar la fase tal que sea un multiplo de 2π.

[11]Es bien sabido que los quarks sí tienen una carga $q = \pm e/3$ ó $q = \pm 2e/3$, por lo que parecen violar la condición de cuantización de carga. Sin embargo QCD, la teoría gauge que describe el comportamiento de los quarks, predice un efecto llamado *libertad asintótica*, responsable del hecho de que los quarks no puedan aparecer libremente. Debido a este *confinamiento de quarks*, éstos se manifiestan siempre en conjuntos de dos o tres, tal que la carga total del conjunto es un multiplo entero de la carga del electrón.

está descrita por dos cartas (parches), el hemisferio norte y el sur, y la función de transición, que nos cuenta cómo encajan los parches, es justo la transformación gauge $\exp(-iq_e q_m \varphi/2\pi)$, definida en el ecuador, donde se solapan los dos parches. Igual que en el caso del efecto Aharonov-Bohm tenemos por lo tanto una función $\Lambda : S^1 \to S^1$, clasificada por el grupo de homotopía $\pi(S^1) = \mathbb{Z}$. El número entero n que aparece en la condición de cuantización de carga (1.111) es básicamente la clase de homotopía de Λ.

Capítulo 2

El Principio de la Relatividad

¿Cómo puede ser que las matemáticas, después de todo un producto del pensamiento humano independiente de la experiencia, son tan admirablemente adecuadas para describir los objetos de la realidad?
(A. Einstein)

2.1. El Principio de la Relatividad

Puede resultar sorprendente, pero históricamente el Principio de la Relatividad es mucho más antiguo que la Teoría de la Relatividad, incluso más antiguo que la mecánica clásica de Newton. Fue formulado por Galileo Galilei (1564 - 1642) alrededor de 1600, como un argumento en la discusión del heliocentrismo versus el geocentrismo. Los defensores del geocentrismo, que en este momento creían firmemente en la física de Aristóteles (384 A.C. - 322 A.C.) y en la astronomía de Ptolomeo (c. 100 - c. 170), argumentaban que, si fuera verdad que la Tierra se moviera alrededor del Sol y alrededor de su eje, ¿por qué no lo notamos? ¿Por qué una bola que dejamos caer desde una torre alta termina al pie de la torre y no a cierta distancia hacia el Oeste, debido a la supuesta rotación de la Tierra de Oeste a Este?

Como respuesta a este argumento Galilei introdujo una nueva idea: la *inercia*. De sus múltiples experimentos con bolas rodando sobre planos inclinados, había llegado a la conclusión de que una masa en movimiento uniforme rectilíneo mantendrá eternamente este movimiento mientras que no actúe ninguna fuerza exterior sobre ella (nótese que esto no es nada menos que la Segunda Ley de Newton, $\vec{F} = m\vec{a}$, en forma cualitativa). Como ejemplo, Galilei dijo que si dejamos caer una bola desde la gavia de un barco en movimiento (uniforme), la bola tocará la cubierta en el pie del mástil y no más hacia la popa, puesto que la bola conserva la velocidad uniforme del barco durante su caída. De la misma manera, el argumento de los geocentristas no demostraba necesariamente que la Tierra esté en reposo.[1]

El propio Galilei se dio cuenta de que la formulación inversa también es válida. Imaginémosnos un observador que se mueve con la misma velocidad que una masa en movimiento uniforme y rectilíneo. Para este observador la masa estará en reposo. Dado que, por el principio de inercia, sin influencias externas ambos mantendrán su movimiento, para el observador esta masa seguirá en reposo hasta que actúe una fuerza exterior sobre ella.

[1]En realidad, si la bola cae desde una altura suficientemente alta, se espera el efecto contrario: antes de caer, el sistema Tierra-torre-bola gira como un cuerpo rígido, de modo que la velocidad inicial de la bola es mayor cuanto más alta la torre. Durante la caída, la bola conservará esa velocidad y acabará por lo tanto una distancia (en la práctica despreciable) más hacia el Este.

Por lo tanto la conclusión que sacó Galilei es que un observador no es capaz de determinar si él está en un sistema que está en reposo o en movimiento uniforme y rectilíneo. El observador puede lanzar o dejar caer masas, dejar rodar bolas sobre planos inclinados, medir el periodo de péndulos, los resultados serán los mismos en movimiento (uniforme) que en reposo. Más general, una persona encerrada en una caja, con todos los experimentos mecánicos imaginables a su disposición, no tiene manera alguna de determinar su estado de movimiento, sin mirar por una ventanilla. Este principio se llama el *Principio de la Relatividad*.

Principio de la Relatividad (formulación de Galilei): *Es imposible determinar a base de experimentos (mecánicos) si un sistema de referencia está en reposo o en movimiento uniforme y rectilíneo.*

Aunque en la vida cotidiana estamos muy acostumbrados a notar si nos movemos o no, hay numerosos ejemplos en los que se aplica claramente el Principio de la Relatividad. Mirando por la ventana de un tren en la estación, no sabemos si empieza a moverse nuestro tren o el de al lado. En un atasco delante de un semáforo pensamos que estamos rodando hacia atrás, si el coche de delante sale. Dos balsas en medio del océano, que se alejan una de otra, no pueden determinar si la primera se aleja de la segunda, o vice versa.

El Principio de la Relatividad, por muy trivial que pueda parecer, es un principio profundísimo y tiene, como veremos en este curso, consecuencias muy lejanas, determinando la forma de las leyes de la física y la estructura del espacio y del tiempo. El Principio de la Relatividad, aunque formulado ya alrededor del año 1600, es la base directa de la teoría de la relatividad especial (1905) y su generalización, el Principio de la Equivalencia (Capítulo 9), la de la teoría de la relatividad general (1916).

Para entender bien su importancia, formularemos el Principio de otra manera. Definimos como *sistemas inerciales* un conjunto de sistemas de referencia que están en reposo o en movimiento uniforme y rectilíneo con respecto a un sistema previamente elegido como inercial.[2] Ahora, el Principio de la Relatividad es solamente válido dentro de la misma clase de sistemas inerciales (ya que salir de la clase de sistemas inerciales implica aceleraciones, que son medibles a través de experimentos). Sin embargo, dentro de una clase de sistemas inerciales, no hay observadores privilegiados: dentro de la misma clase de sistemas inerciales no se puede determinar qué sistema está en "reposo absoluto" y cuál en "movimiento absoluto". Todos los experimentos dan el mismo resultado para cualquier observador y por lo tanto todos los observadores inerciales ven la misma física. Ninguno de ellos tiene un punto de vista priveligiado frente a los otros.

Principio de la Relatividad (formulación de equivalencia): *Todos los sistemas inerciales son equivalentes, es decir, todos los observadores inerciales ven la misma física.*

Nadie se había dado cuenta realmente de la importancia de esta formulación hasta finales del siglo XIX, cuando Henri Poincaré (1854-1912) lo tomó como principio básico y sugirió que hacía falta una nueva formulación de la mecánica, basada en este principio. Aunque Poincaré nunca llegó a dudar de la existencia del éter, sí se hacía serias preguntas sobre el significado de simultaniedad de eventos y derivaba la regla relativista de la suma de velocidades. Sin embargo, no llegó a formular una mecánica relativista completa. Este logro está a nombre de Albert Einstein (1879-1955). El razonamiento de Einstein era que, si todos los sistemas inerciales son equivalentes y todos los observadores inerciales ven la misma física, entonces todos estos observadores inerciales deben llegar a las mismas leyes de la física, si quieren apuntar los resultados de sus experimentos. En otras palabras:

[2] Nótese que esta definición es altamente circular y no precisa cuáles son los observadores inerciales. Sólo dice que si conocemos uno, podemos encontrar todos de su clase. Veremos en la sección 9.6 por qué es tan difícil identificar observadores como inerciales, pero ya podemos adelantar que el problema toca la esencia de la teoría de la relatividad general.

Principio de la Relatividad (formulación de Einstein): *Las leyes de la física deben tener la misma forma en todos los sistemas inerciales.*

El gran logro de Poincaré y Einstein fue darse cuenta dde que el Principio de la Relatividad no sólo se aplica a la mecánica, sino a la física entera, incluidos el electromagnetismo y la óptica. Esta extensión es importante, puesto que como veremos en la sección 2.3, hay un conflicto directo entre la mecánica newtoniana y el electromagnetismo de Maxwell (1831-1879) si aplicamos el Principio de la Relatividad.

Formulado de esta manera, el Principio de la Relatividad impone ciertas condiciones sobre la forma de las leyes de la física. Existen unas transformaciones, llamadas *cambios de coordenadas*, que relacionan la posición, velocidad, etc. de un observador con otro. Para no salir de la clase de sistemas inerciales, estos cambios de coordenadas tienen una forma específica y estas transformaciones tienen la estructura matemática de un grupo. La formulación de Einstein del Principio de la Relatividad implica por lo tanto, que las leyes de la física tienen que ser de tal forma que, al hacer un cambio de coordenadas entre dos sistemas inerciales, tienen que tener la misma forma. En otras palabras: las leyes de la física deben ser invariantes y las cantidades físicas que aparecen en estas leyes tienen que ser tales que transformen bien bajo las transformaciones de ese grupo.

Principio de la Relatividad (formulación covariante): *Las leyes de la física transforman bien bajo las transformaciones del grupo de simetría que relacionan a los distintos observadores.*

Investiguemos ahora en detalle cuáles son estas transformaciones de simetría y cómo las leyes conocidas transforman bajo ellas.

2.2. Simetrías y la forma de las leyes de la física

Las leyes de la física son invariantes bajo traslaciones en el tiempo

$$t \to t + \Delta t, \tag{2.1}$$

para cualquier valor de Δt. En otras palabras, no hay un momento especial $t = 0$ y las leyes son válidas en todos los momentos t. Esta simetría encaja bien con la hipótesis básica de la ciencia de que los resultados son reproducibles,[3] por la sencilla razón de que si no la teoría no sería refutable. Si un científico obtiene cierto resultado en el momento $t = t_1$, otro investigador debería ser capaz de obtener el mismo resultado a $t = t_1 + \Delta t$. La simetría (2.1) se llama la *homogeneidad del tiempo*.

De la misma manera, las leyes de la física son invariantes bajo traslaciones en el espacio

$$\vec{r} \to \vec{r} + \vec{r}_0, \tag{2.2}$$

para cualquier vector $\vec{r}_0$. Físicamente esto corresponde al hecho de que no importa el lugar donde hacemos los experimentos y matemáticamente a que no hay un punto especial en el espacio y que podemos elegir el origen del sistema de referencia donde queramos. Esta simetría se llama la *homogeneidad del espacio*.

La *isotropía del espacio* es el hecho de que todas las direcciones son equivalentes y que no hay ninguna dirección preferida. En otras palabras, la orientación de un experimento físico es

[3] Aunque existe un *Annals of Improbable Research*, editado por la misma gente que cada año entrega *the Ignoble Prizes*, un premio por investigaciones *"that cannot or should not be reproduced."*.

irrelevante. La isotropía implica una invariancia bajo rotaciones ortogonales arbitrarias

$$x'^i = \sum_{j=1}^{3} M^i{}_j x^j , \tag{2.3}$$

donde $M^i{}_j$ es la matriz ortogonal que representa una rotación $SO(3)$.

Como ya mencionamos antes, estas simetrías imponen restricciones en la forma que pueden tener las leyes físicas. La homogeneidad del tiempo causa que un lagrangiano[4] $\mathcal{L} = T - V$, que describe un sistema de objetos en posiciones q_i con velocidades $\dot{q}_i$, no pueda depender explícitamente de t:

$$\partial_t \mathcal{L}(q_i(t), \dot{q}_i(t)) = 0. \tag{2.4}$$

Como indicado, las variables q_i y $\dot{q}_i$ sí pueden ser dinámicas y depender de t, pero la dependencia del lagrangiano de t sólo puede darse a través de éstas variables. La homogeneidad del espacio implica que las interacciones permitidas en el potencial $V(q_i)$ sólo pueden depender de las distancias entre dos partículas, pero no de las posiciones de las partículas con respecto a cierto origen. Y finalmente la isotropía del espacio implica que el potencial V tiene que ser un potencial central, que depende de la distancia radial r, pero no de los ángulos.

Efectivamente, los potenciales que uno encuentra en teorías de gravedad, electromagnetismo, física molecular o nuclear son típicamente de la forma $V = V(|\vec{r}_1 - \vec{r}_2|)$, que claramente satisface los requisitos mencionados arriba.

Las simetrías no sólo restringen la forma que puede tener una teoría, sino también ayudan a la hora de encontrar soluciones de las ecuaciones de movimiento. Si, por ejemplo, queremos encontrar la expresión para el campo eléctrico, causado por un electrón en reposo en el origen de un sistema de coordenadas, tenemos que resolver la *ecuación de Laplace* (en coordenadas esféricas) para el potencial ϕ,

$$\nabla^2\phi \;\equiv\; \frac{1}{r^2}\,\partial_r(r^2\partial_r\phi) \;+\; \frac{1}{r^2\sin\theta}\,\partial_\theta(\sin\theta\partial_\theta\phi) \;+\; \frac{1}{r^2\sin^2\theta}\,\partial^2_\varphi\phi \;=\; 0. \tag{2.5}$$

Para resolver la ecuación (2.5), se podría aplicar el método de separación de variables y buscar la solución más general en función de las funciones armónicas esféricas $Y_{lm}(\theta,\varphi)$, pero nos ahorraremos mucho trabajo si utilizamos la simetría del sistema. Dado que el electrón en el origen posee una simetría esférica, el campo eléctrico también tiene que reflejar esta simetría: la solución dependerá sólo de la coordenada radial r y no de los ángulos θ y φ. La ecuación diferencial parcial (2.5) se reduce por lo tanto a una ecuación diferencial ordinaria en r y es fácil averiguar que la solución para el campo eléctrico es la solución conocida de la electrostática:

$$\phi(r) = \frac{Q}{r} \qquad\Longrightarrow\qquad \vec{E} = -\vec{\nabla}\phi = -\frac{Q}{r^2}\,\vec{e}_r. \tag{2.6}$$

Aparte de las simetrías bajo traslaciones y rotaciones de arriba, el Principio de la Relatividad nos proporciona otra simetría, relacionada con observadores en movimiento relativo. Ya que el Principio de la Relatividad impone que un observador no es capaz de medir la velocidad absoluta de su sistema de referencia, sino solamente velocidades relativas entre sistemas de referencia, obviamente esto implica que las leyes de la física no pueden ser formuladas en términos de velocidades. Sólo cambios de velocidad son admisibles, ya que estos son independientes de los obervadores. Efectivamente, las leyes de Newton están formuladas en función de la aceleración $\vec{a}$, y no de la velocidad $\vec{v}$.

[4]A partir de ahora consideraré los conceptos *lagrangiano* y *teoría* como sinónimos: cada lagrangiano, al ser la integral de las ecuaciones de movimiento, define los grados de libertad y dinámica en cuestión y por lo tanto la teoría entera. Vice versa consideraré que una teoría está bien definida si existe un lagrangiano que describe su dinámica. (No consideraremos en este libro los casos donde un conjunto de ecuaciones de movimiento no son integrables a una acción.)

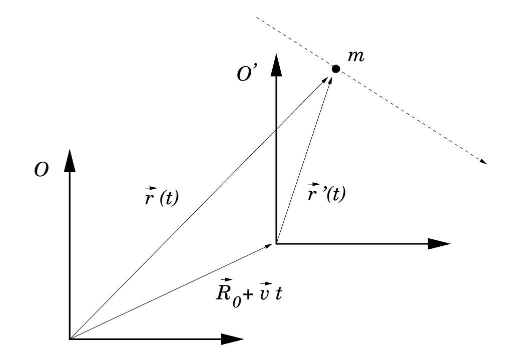


Figura 2.1: *Las transformaciones de Galilei: Un observador $\mathcal{O}'$ se mueve con velocidad $\vec{v}$ con respecto a un observador $\mathcal{O}$, tal que la posición de $\mathcal{O}'$ en las coordenadas de $\mathcal{O}$ es $\vec{R}_0 + \vec{v}t$. Una partícula, que tiene posición $\vec{r}\,'(t)$ en las coordenadas de $\mathcal{O}'$ tendrá posición $\vec{r}(t) = \vec{R}_0 + \vec{r}\,'(t) + \vec{v}t$ en las coordenadas de $\mathcal{O}$.*

Pero hay más. Si dos observadores observan el mismo suceso, cada uno en su propio sistema de referencia, tiene que haber una manera de relacionar los resultados de un observador con los resultados del otro. La relación entre los resultados de diferentes observadores se llama *cambio de coordenadas*. Las rotaciones y las traslaciones en el tiempo y el espacio son ejemplos de cambios de coordenadas, pero por lo menos igual de importantes son los cambios entre dos observadores que están en movimiento uniforme rectilíneo relativo.

Consideremos por simplicidad dos observadores $\mathcal{O}$ y $\mathcal{O}'$ que tienen sus sistemas de coordenadas orientados de tal manera que los ejes de sus coordenadas cartesianas son paralelos. El observador $\mathcal{O}'$ se mueve con una velocidad constante $\vec{v}$ con respecto al sistema de referencia de $\mathcal{O}$, así que la posición del orígen de $\mathcal{O}'$ con respecto a $\mathcal{O}$ en un momento t arbitrario viene dado por $\vec{R}_0 + \vec{v}t$, donde $\vec{R}_0$ es la posición de $\mathcal{O}'$ a $t = 0$ en el sistema de $\mathcal{O}$ (véase Figura 2.1). Una masa m que se encuentra en el momento t en la posición $\vec{r}\,'$ en el sistema de coordenadas de $\mathcal{O}'$, tendrá para $\mathcal{O}$, por la regla de suma de vectores, la posición

$$\vec{r} = \vec{R}_0 \ + \vec{r}\,' + \ \vec{v}t. \tag{2.7}$$

Los observadores $\mathcal{O}$ y $\mathcal{O}'$ podrán relacionar sus mediciones de posición a través de estas transformaciones, llamadas las *transformaciones de Galilei*. Muchas veces se presentan para el caso especial cuando los dos sistemas de referencia coinciden a $t = 0$ y $\mathcal{O}'$ se mueve a lo largo del eje x positivo de $\mathcal{O}$ (es decir $\vec{R}_0 = 0$ y $\vec{v} = v\vec{e}_x$). En este caso (2.7) se reduce a

$$x = x' + vt, \qquad y = y', \qquad z = z'. \tag{2.8}$$

Si la masa m se mueve con velocidad $\vec{V}$ para $\mathcal{O}$ y con velocidad $\vec{V}'$ para $\mathcal{O}'$, su trayectoria viene dada por $\vec{r}(t) = \vec{r}_0 + \vec{V}t$ en las coordenadas de $\mathcal{O}$ y por $\vec{r}\,'(t) = \vec{r}_0{}' + \vec{V}'t$ en coordenadas de $\mathcal{O}'$, donde $\vec{r}_0$ y $\vec{r}_0{}'$ son las posiciones de la masa en el momento $t = 0$ en las coordenadas de $\mathcal{O}$ y de $\mathcal{O}'$ respectivamente. Utilizando (2.7), vemos que las dos trayectorias están relacionadas a través de

$$\vec{r}_0 \ + \ \vec{V}t \ = \ \vec{R}_0 \ + \ \vec{v}t \ + \ \vec{r}_0{}' \ + \ \vec{V}'t. \tag{2.9}$$

Esta relación obviamento no es más que una reformulación de (2.7) para t arbitrario, pero su interés está en que nos permite encontrar una relación entre las velocidades $\vec{V}$ y $\vec{V}'$ medidas por $\mathcal{O}$ y $\mathcal{O}'$. Derivando (2.9) con respecto al tiempo, obtenemos que la velocidad que $\mathcal{O}$ mide para la masa m es la suma de la velocidad que mide $\mathcal{O}'$ y la velocidad $\vec{v}$ de $\mathcal{O}'$ relativa a $\mathcal{O}$:

$$\vec{V} = \vec{v} + \vec{V}'. \tag{2.10}$$

Esta regla se llama la *regla (clásica) de la suma de las velocidades* y parece confirmar lo que conocemos intuitivamente de la vida cotidiana. Derivando (2.10) otra vez con respecto a t, vemos que

$$\vec{a} = \vec{a}\,'. \tag{2.11}$$

En otras palabras, (2.7), (2.10) y (2.11) nos confirman los resultados conocidos de que en la mecánica newtoniana las posiciones y las velocidades son relativas, pero las aceleraciones son absolutas.

Como hemos visto, el Principio de la Relatividad dice que diferentes observadores inerciales ven la misma física. Así que las leyes de la física que estos observadores formulan tienen que ser invariantes bajo la transformación (2.8) (o en general (2.7)). Las transformaciones de Galilei, junto con las rotaciones (2.3) y las traslaciones espaciales (2.2) y temporales (2.1) forman un grupo (ejerc.), llamado el *grupo de Galilei*. Como veremos en la siguiente sección, el grupo de Galilei es el grupo de simetrías de la mecánica newtoniana.

2.3. Invariancia de la mecánica newtoniana y las leyes de Maxwell

Investigaremos la invariancia de la segunda ley de Newton en el caso en que las fuerzas sean conservativas, es decir que son derivables de un potencial $\vec{F} = -\vec{\nabla}V$. Esto es el caso para todas las interacciones fundamentales, ya que los efectos no-conservativos, como el rozamiento, son una descripción efectiva de interacciones conservativas complicadas a nivel molecular.

En componentes la segunda ley de Newton tiene la forma

$$-\frac{\partial V}{\partial x_i} = m\ddot{x}^i. \tag{2.12}$$

A base de conservación de momento, se puede demostrar que la masa m de una partícula tiene el mismo valor para todos los observadores. En otras palabras, la masa es *invariante* bajo cambios de coordenadas, es decir, que toma el mismo valor para todos los observadores. También hemos visto que los potenciales típicos de gravedad, electromagnetismo y física molecular, atómico o nuclear son de la forma $V = V(|\vec{r}_1 - \vec{r}_2|)$, tal que estos potenciales son invariantes bajo rotaciones y traslaciones.[5]

En el lado derecho de (2.12) aparece la segunda derivada de la posición $\ddot{x}^i$, de forma que con la regla de la cadena es fácil ver que la ecuación es invariante bajo las transformaciones (2.1), (2.2) y (2.7). En particular, el hecho de que la segunda ley de Newton venga en función de la aceleración y no de la velocidad es una consecuencia directa del Principio de la Relatividad: si aparecieran velocidades, habría que especificar con respecto a qué sistema de referencia, mientras las aceleraciones son absolutas en la mecánica newtoniana.

La invariancia bajo rotaciones es un poco más sutil: derivando dos veces la transformación (2.3) con respecto al tiempo, está claro que la aceleración transforma como

$$\ddot{x}'^i = \sum_{j=1}^{3} M^i{}_j \ddot{x}^j. \tag{2.13}$$

Aunque, por ser un potencial central, V es invariante bajo rotaciones, su gradiente no lo es, debido a la transformación de las derivadas. Aplicando la regla de la cadena tenemos (o véase Capítulo 4)

$$\frac{\partial V}{\partial x'_i} = \sum_{j=1}^{3} M^i{}_j \frac{\partial V}{\partial x_j}. \tag{2.14}$$

[5] Ojo, el hecho de que V sea invariante bajo rotaciones no significa que $\vec{\nabla}V$ lo sea. Es más, en seguida demostraremos que $\vec{\nabla}V$ no es invariante, sino que transforma de manera covariante.

Combinando (2.13) y (2.14), obtenemos que si para un observador es válida la segunda ley (2.12), otro observador relacionado con el primero a través de una rotación (2.3) verá la segunda ley como

$$-\frac{\partial V}{\partial x'_i} = m\ddot{x}'^i. \tag{2.15}$$

Las ecuaciones (2.13) y (2.14) dicen que los dos observadores obtendrán resultados diferentes para las componentes de la fuerza y la aceleración, pero cada uno medirá, componente por componente, que la fuerza es proporcional a la aceleración, con la masa como constante de proporcionalidad. Si la ley de Newton es válida para un observador, también lo es para el otro. Si objetos (y leyes) transforman de esta manera decimos que no son invariantes sino *covariantes*.

La razón por la que esto funciona es que tanto la fuerza $-\vec{\nabla}V$ como la aceleración $\ddot{\vec{x}}$ transforman de la misma manera bajo rotaciones. Ambas son vectores en $\mathbb{R}^3$ y transforman como tales bajo el grupo $O(3)$ de rotaciones en $\mathbb{R}^3$. La lección que aprendemos de esto es que para que las leyes de la física sean invariantes o covariantes, hay que expresarlas en función de objetos que transformen bien bajo el grupo de simetría, en este caso vectores. Si no, estas leyes sólo tendrían sentido en un sistema de referencia específico.

Con estos cálculos hemos demostrado la invariancia y covariancia de la mecánica newtoniana bajo el grupo de Galilei. El problema aparece cuando intentamos averiguar la invariancia del otro pilar de la física clásica: el electromagnetismo en forma de la teoría de Maxwell. Las leyes de Maxwell vienen dadas (en unidades de Lorentz-Heaviside) por

$$\begin{aligned} &\vec{\nabla}\cdot\vec{E} = \rho, && \vec{\nabla}\cdot\vec{B} = 0, \\ &\vec{\nabla}\times\vec{E} = -\frac{1}{c}\,\partial_t\vec{B}, && \vec{\nabla}\times\vec{B} = \frac{1}{c}\Big(\vec{\jmath} + \partial_t\vec{E}\Big), \end{aligned} \tag{2.16}$$

donde $\vec{E}$ y $\vec{B}$ son los campos eléctricos y magnéticos respectivamente, ρ la densidad de carga, $\vec{\jmath}$ la densidad de corriente y $c = 299792,458$ km/s la velocidad de la luz. La velocidad de la luz aparece porque la teoría de Maxwell predice la existencia de ondas electromagnéticas (o sea luz), que viajan con la velocidad c.

El hecho de que aparezca explícitamente la velocidad de la luz es extraño porque parece violar el Principio de la Relatividad: c es la velocidad de la luz ¿con respecto a qué observador? La interpretación generalmente aceptada a finales del siglo XIX era que las ondas electromagnéticas necesitaban un medio para propagarse, el *éter*, y c sería entonces la velocidad de la luz con respecto a ese éter.

Un observador que se mueve con una velocidad v con respecto al éter mediría entonces, por la regla de la sumatoria de las velocidades (2.10), que la luz se propaga con una velocidad

$$c' = c \pm v, \tag{2.17}$$

dependiendo de si se acerca o se aleja de la fuente de luz. En particular, un observador que se mueve a la velocidad de la luz en la misma dirección que una onda electromagnética, vería esa onda como estática ($c' = 0$), ya que estaría en reposo con respecto al observador. Sin embargo, las ecuaciones de Maxwell en el vacio ($\rho = \vec{\jmath} = 0$) no permiten soluciones de campos estáticos, al menos que sean constantes (ejerc.). Por lo tanto, ese observador vería una perfil del campo electromagnético que no está permitido por las ecuaciones de Maxwell, lo que implica un grave problema teórico.

Pero también hay graves problemas experimentales: entre 1881 y 1930 una serie de experimentos (de los cuales el de los americanos Albert Michelson (1852 - 1931) y Edward Morley (1838 - 1923) de 1887 es el más famoso) intentaron determinar la velocidad v de la Tierra con respecto al éter, midiendo con mucha precisión la velocidad de la luz (2.17) con respecto a la Tierra. Sin embargo a pesar de la gran precisión, los resultados nunca mostraron (dentro del margen de error)

ninguna diferencia entre la velocidad de la luz con respecto a la Tierra y la velocidad con respecto al éter. En 1907 Albert Michelson recibió el Premio Nobel, no por el resultado de su experimento, sino por inventar el interferómetro con que había hecho el experimento.

Para explicar los resultados experimentales de Michelson y Morley, el físico irlandés George FitzGerald (1851 - 1901) e indepedientemente el holandés Hendrik Anton Lorentz (1853-1928) propusieron que cualquier objeto en movimiento con respecto al éter se contrae con un factor $1/\sqrt{1-v^2/c^2}$ en la dirección del movimiento. Aunque a estas alturas la suposición era completamente ad hoc, sí explicaba los resultados correctos, ya se supuestamente los brazos del interferómetro se contraen de tal modo que cualquier medición de la velocidad de la luz siempre daría como valor c.

Lorentz iba aún más lejos, dándose también cuenta de que las leyes de Maxwell no transforman de manera covariante bajo las transformaciones de Galilei (2.7), sino bajo las transformaciones

$$t' = \frac{t - vx/c^2}{\sqrt{1-v^2/c^2}}, \qquad x' = \frac{x - vt}{\sqrt{1-v^2/c^2}}, \tag{2.18}$$

si asumimos que las cargas y las corrientes transforman como

$$\rho' = \frac{\rho - vj_x/c^2}{\sqrt{1-v^2/c^2}}, \qquad j'_x = \frac{j_x - v\rho}{\sqrt{1-v^2/c^2}}, \tag{2.19}$$

y los campos $\vec{E}$ y $\vec{B}$ como

$$\begin{aligned} E'_x &= E_x, & E'_y &= \frac{E_y - vB_z/c}{\sqrt{1-v^2/c^2}}, & E'_z &= \frac{E_z + vB_y/c}{\sqrt{1-v^2/c^2}}, \\ B'_x &= B_x, & B'_y &= \frac{B_y + vE_z/c}{\sqrt{1-v^2/c^2}}, & B'_z &= \frac{B_z - vE_y/c}{\sqrt{1-v^2/c^2}} \end{aligned} \tag{2.20}$$

(ejerc.). La transformación (2.18) se llama la *transformación de Lorentz* y difiere de la transformación de Galilei (2.7) por el factor $1/\sqrt{1-v^2/c^2}$, que es despreciable cuando $v \ll c$, pero muy grande cuando v tiende a c. Igual que las transformaciones de Galilei, las transformaciones de Lorentz forman un grupo (ejerc.), llamado el *grupo de Lorentz*.

Parece por lo tanto que tenemos dos grupos de simetría distintos en la Naturaleza: el grupo de Galilei para la mecánica newtoniana y el grupo de Lorentz para el electromagnetismo. Hay tres posibles actitudes ante esta situación peculiar:

1. Abandonar el Principio de la Relatividad: los sistemas inerciales no son todos equivalentes, sino el sistema que está en reposo con respecto al éter es especial, puesto que sólo aquí las leyes de Maxwell son válidas.

2. Modificar las leyes de Maxwell, tal que transformen bien bajo las transformaciones de Galilei.

3. Modificar la mecánica newtoniana, tal que transforme bajo las transformaciones de Lorentz en lugar de las de Galilei.

La primera postura es la que tomó Lorentz. Él creía que su contracción de Lorentz era un hecho objetivo, absoluto frente al sistema del éter. En cuanto a la segunda postura, ha habido intentos de por ejemplo Walter Ritz (1878 - 1909) e incluso Einstein en cierto momento, de derivar leyes electromagnéticas que transformen bajo el grupo de Galilei, pero estas llevan a predicciones que fueron experimentalmente refutadas.

La tercera opción es, como sabemos, la que eligió Einstein en 1905 y es sin duda la más valiente. Fíjese que en 1905 la mecánica newtoniana llevaba funcionando con gran éxito ya más de dos siglos, mientras la teoría de Maxwell apenas tenía 40 años y su gran predicción, la existencia de ondas electromagnéticas, sólo había sido confirmada por Hertz alrededor de 1890.

Einstein empieza su famoso artículo *Sobre la electrodinámica de los cuerpos en movimiento*, con una observación que expresa su descontento con la intepretación tradicional de algunos fenómenos electromagnéticos: cuando un imán se acerca a un aro conductor, la ley la inducción magnética de Faraday explica la aparición de un campo eléctrico circular que induce una corriente en el conductor. Sin embargo si el conductor se acerca al imán, no hay ningún campo eléctrico, pero la fuerza de Lorentz $\vec{F} = q(\vec{E} + \vec{v} \times \vec{B}/c)$ actúa sobre las cargas libres en el conductor, generando una corriente. A Einstein le parecía raro estas dos explicaciones distintas, si en el fondo se trataba del mismo fenómeno, visto desde puntos de vistas diferentes, y argumentaba la necesidad de una descripción independiente del estado de movimiento del observador.

El resultado de la valiente postura de Einstein es lo que ahora conocemos como la teoría de la relatividad especial y aparte de llevar a una mecánica relativista, también llevó a un concepto nuevo de espacio y tiempo.

2.4. El Principio de la Relatividad y el espacio absoluto

Es curioso, desde nuestro punto de vista moderno post-relativista, que el Principio de la Relatividad, que es tan importante en el desarollo de la teoría de la relatividad, no juegue ningún papel de importancia en la mecánica newtoniana. Se mencionan las transformaciones de Galilei (2.8) en cualquier curso de física general, aunque más bien de manera anecdótica, puesto que luego no se usan en ningún lado. Se podría aprender y entender toda la física newtoniana igual de bien, sin conocer la simetría (2.8).

Gran parte de la responsabilidad de esto la tiene Isaac Newton (1642 - 1727), por el énfasis que hizo sobre el espacio, el tiempo y el movimiento absolutos. En el *Principia* escribe

> "El tiempo absoluto, verdadero y matemático [...] fluye igualmente sin referencia a algo externo."
> "El espacio absoluto, en su propia naturaleza, sin referencia a algo externo, permanece siempre similar e inmovible."
> "[E]n un barco de vela, la posición relativa de un cuerpo es aquella parte del barco [...] que el cuerpo ocupa, y que por lo tanto se mueve junto con el barco. El reposo relativo es la continuación del cuerpo en alguna parte del barco [...]. Pero reposo verdadero, absoluto es la continuidad del cuerpo en alguna parte del espacio absoluto."

Aunque Newton no necesitaba recurrir a estas ideas tan absolutas para formular su mecánica, lo hizo porque no le gustaba el relativismo de René Descartes (1596 - 1650), que le olía a ateísmo. Algunos de sus contemporáneos, como Gottfried Leibniz (1646 - 1716), han criticado mucho estas ideas absolutas de Newton, pero según iba aumentando el éxito del *Principia* y la autoridad de Newton, la discusión se trasladaba cada vez más a un plano filosófico,[6] mientras que el significado físico se cuestionaba cada vez menos. Veremos que un proceso similar ha ocurrido con la idea de la gravedad como acción a distancia, concepto cuestionado por el propio Newton y sus contemporáneos, pero generalmente aceptada por generaciones posteriores.

A finales del siglo XVIII, la autoridad del *Principia* había impuesto una manera de pensar y hacer física, de modo que el Principio de la Relatividad quedaba confinado a un rincón olvidado

[6] Por ejemplo, los empiristas como George Berkeley (1685 - 1753) y David Hume (1711 - 1776) rechazaban las ideas del espacio y el tiempo absolutos a base de que no eran observables con los sentidos, mientras que para Immanuel Kant (1724 - 1804) eran ejemplos estupendos de sus *categorías*, conceptos innatos en el conocimiento humano.

de la mecánica. Habría que esperar hasta el físico-filósofo austríaco Ernst Mach (1838 - 1916) a finales del siglo XIX para una primera crítica severa a las ideas absolutas de Newton. Como los filósofos empiristas, Mach creía que la única fuente de conocimiento son las impresiones recibidas por los sentidos. Para él la ciencia era sólo un conjunto de hechos y datos sacados de experimentos y observaciones, y las leyes de la física, como la del gas ideal o de la mecánica newtoniana, no son más que maneras económicas de resumir experiencia. Sus ideas le hacían rechazar ideas teóricas por no ser observables, como el espacio, el tiempo y el movimiento absoluto, pero también los átomos, cuya existencia nunca aceptó.

Pero el gran mérito de rescatar el Principio de la Relatividad corresponde, como ya dijimos en la sección 2.1, a Poincaré y a Einstein, que lo declararon como *primer principio*, una idea cuya validez no se puede demostrar directamente, sino incorporándola en una teoría. Sin embargo, un primer principio es más básico que una teoría, ya que es independiente y previo a la teoría y puede ser implementado en teorías de varias maneras. Lo que nos enseña la historia de la teoría de la relatividad es que, si hay un conflicto entre una teoría concreta y un primer principio (en este caso la mecánica newtoniana y el Principio de la Relatividad), en general de lo que uno se puede fiar es del primer principio.

Lo que hizo Einstein fue introducir una nueva manera de hacer física teórica. Hasta entonces un artículo teórico típicamente empezaba resumiendo unos datos experimentales que no estaban de acuerdo con las teorías del momento y seguía proponiendo un nuevo modelo que explicaba las discrepancias. El artículo de Einstein sobre la relatividad especial, sin embargo, empezaba con la observación de que tradicionalmente se interpreta la interacción entre una partícula cargada y un campo magnético de distinta manera, dependiendo de si se mueve la partícula o el imán, lo que está en contradicción con el Principio de la Relatividad. A continuación introdujo unos postulados básicos, como primeros principios y derivó de allí toda una nueva mecánica y unos nuevos conceptos de espacio y tiempo. Sólo al final sugirió dos maneras de verificar la nueva teoría experimentalmente: el efecto Doppler y la aberración relativista. Esta manera de hacer física teórica hoy en día todavía está considerada entre los físicos teóricos una manera muy elegante que se imita en numerosas ocasiones.

Capítulo 3

Breve repaso de la relatividad especial

La relatividad nos enseña la conexión entre las diferentes descripciones de la misma realidad.
(A. Einstein)

3.1. La relatividad del tiempo y el espacio

Toda la teoría de la relatividad especial está basada en dos postulados:

1. Todas las leyes de la física son válidas para todos los sistemas inerciales.

2. La velocidad de la luz en el vacío es igual para todos los observadores y tiene el valor de 299.792,458 km/s, independientemente del estado de movimiento de la fuente.

El primer postulado es el Principio de la Relatividad que vimos en el capítulo anterior y que fue formulado ya por Galilei para la mecánica alrededor de 1600. Einstein lo amplió para todas las leyes físicas, especialmente para el electromagnetismo, ya que, como vimos, la teoría de Maxwell parecía comportarse de manera diferente a la mecánica newtoniana. El Principio de la Relatividad afirma que no existe ningún experimento físico que sea capaz de distinguir si un observador está en reposo o en movimiento uniforme rectilíneo. Dos observadores que se mueven uno con respecto al otro ven la misma física y, por lo tanto, las leyes de la física deben escribirse de modo que no cambien al pasar de un sistema de referencia a otro.

El segundo postulado es más sorprendente, pero hay motivos teóricos y experimentales para creerlo. Los motivos experimentales son los experimentos fracasados de Michelson y Morley (1881 y 1887), que querían medir la velocidad de la Tierra con respecto al éter: suponiendo que la velocidad de la luz con respecto al éter es c y la de la Tierra alrededor del Sol es v, la velocidad de la luz medida en la Tierra debería ser $c - v$ o $c + v$, dependiendo de si se mide en la dirección de movimiento de la Tierra, o la dirección opuesta. Los resultados experimentales sin embargo no lograron medir ninguna diferencia de velocidad, obteniendo siempre el mismo valor para c. El motivo teórico es la teoría de Maxwell que afirma la existencia de ondas electromagnéticas (luz), cuya velocidad c es una constante predicha por la teoría. Si se asume que todas las leyes de la física son válidas para todos los observadores, hay que aceptar que c es una constante universal.

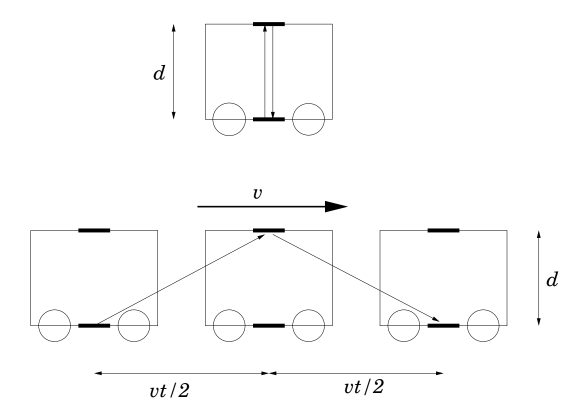


Figura 3.1: *El reloj de luz en un tren en movimiento visto por un observador $\mathcal{O}'$ dentro del tren (arriba) y un observador $\mathcal{O}$ en el andén (abajo).*

Llevar estos dos postulados hasta sus últimas consecuencias implica abandonar las ideas intuitivas del espacio y el tiempo. Investigaremos las implicaciones con una serie de experimentos ficticios.

Considérese un tren que se mueve con velocidad v en movimiento uniforme rectilíneo con respecto al andén de una estación. El pasajero en el tren $\mathcal{O}'$ dispone de un reloj de luz, que consiste en dos espejos colocados uno encima del otro a una altura d y un pulso de luz que viaja continuamente entre los dos espejos. Por lo tanto, $\mathcal{O}'$ medirá que el tiempo Δt que tarda la luz en subir y bajar entre los dos espejos es

$$\Delta t' = \frac{2d}{c}. \tag{3.1}$$

Un observador $\mathcal{O}$ en el andén verá este mismo fenómeno de manera distinta: para él la luz sale del espejo de abajo, pero llega al espejo de arriba después de un tiempo $\Delta t/2$ cuando el tren se ha desplazado una distancia $v\Delta t/2$ y otra vez al espejo de abajo después de un tiempo total Δt cuando el tren se ha desplazado una distancia total $v\Delta t$ (véase Figura 3.1). Para $\mathcal{O}$, la luz recorre una trayectoria más larga y, dado que la velocidad de la luz es la misma que para el pasajero, habrá pasado más tiempo entre que la luz saliese y llegase otra vez al espejo de abajo. Concretamente, la distancia que recorre la luz al subir es, por el teorema de Pythagoras

$$\Big(\frac{c\Delta t}{2}\Big)^2 = \Big(\frac{v\Delta t}{2}\Big)^2 + d^2, \tag{3.2}$$

de donde podemos despejar Δt como

$$\Delta t = \frac{2d/c}{\sqrt{1 - v^2/c^2}} = \frac{\Delta t'}{\sqrt{1 - v^2/c^2}}, \tag{3.3}$$

donde en la última igualdad hemos utilizado (3.1) para relacionar el intervalo Δt con el intervalo $\Delta t'$ medido por el pasajero. Vemos por lo tanto que el intervalo de tiempo efectivamente ha sido más largo para el observador en el andén que para el pasajero. Este efecto se llama la *dilatación del tiempo* y es completamente general: los relojes en movimiento corren menos rápido que los relojes en reposo. El factor de corrección

$$\gamma = \frac{1}{\sqrt{1 - v^2/c^2}} \tag{3.4}$$

es una función creciente de v, que siempre es mayor que 1. Obsérvese sin embargo que para velocidades $v \ll c$ mucho más pequeñas que la velocidad de la luz, tenemos que $\gamma \approx 1$, de modo que estos efectos relativistas son completamente despreciables en la vida cotidiana.

Nótese que para obtener la fórmula (3.3), hemos tenido que asumir que $v < c$. La velocidad de la luz surge por lo tanto en la relatividad especial como un límite máximo, que ningún observador, ni ninguna señal puede superar.

Finalmente, es imprescindible darse cuenta de que la situación descrita arriba es completamente simétrica entre los dos observadores. El pasajero $\mathcal{O}'$ en el tren tiene pleno derecho de suponer que él está en reposo, mientras el observador $\mathcal{O}$ en el andén se está moviendo. Repitiendo el mismo argumento, está claro que $\mathcal{O}'$ verá el reloj de luz de $\mathcal{O}$ en el andén avanzar menos rápido que el suyo, dado que para él su propio reloj está en reposo (ejerc.).

Concluimos por lo tanto que cada uno ve el reloj del otro ir más lento que el suyo, puesto que cada uno ve al otro en movimiento. No hay manera de saber cual de los dos "realmente" va más lento, ya que esto depende del punto de vista del observador. Por muy contraintuitiva que pueda parecer, esta situación no lleva a contradicciones, si se toma en cuenta el hecho de que también las distancias son relativas.

Asumimos que el andén tiene una longitud L medida por $\mathcal{O}$. Este puede escribir $L = v\Delta t$, donde Δt es el intervalo que necesita el tren para recorrer el andén. Por otro lado, $\mathcal{O}'$ mide una longitud $L' = v\Delta t'$, donde ahora $\Delta t'$ es el tiempo entre que el pasajero pasa por el principio y el final del andén, medido por él.

Dado que los dos observadores no coinciden en cuánto ha durado el intervalo de tiempo, tampoco se pondrán de acuerdo sobre la longitud del andén. Para $\mathcal{O}'$, el andén mide

$$L' \;=\; v\Delta t' \;=\; v\Delta t\sqrt{1 - v^2/c^2} \;=\; L\sqrt{1 - v^2/c^2}, \tag{3.5}$$

es decir una factor $\sqrt{1 - v^2/c^2}$ más corto que para $\mathcal{O}$. Este efecto se conoce bajo el nombre de *contracción de Lorentz*: los objetos en movimiento sufren una contracción longitudinal con un factor γ^{-1}. Otra vez la situación es simétrica: $\mathcal{O}$ verá el tren contraído con respecto a la medición de $\mathcal{O}'$ (ejerc.).

La contracción de Lorentz y la dilatación del tiempo conspiran para que el conjunto sea consistente. Esto se puede ver en el famoso experimento de los muones. Los muones son partículas elementales con una vida media de 2,2 $\cdot\, 10^{-6}s$ (en reposo), que se forman a unos 15 km de altura en las colisiones de rayos cósmicos con átomos de la atmósfera. Los muones producidos en estas colisiones se mueven típicamente con velocidades $v \sim 0,99\, c$.

Calculando ingenuamente la distancia que recorrerían los muones antes de desintegrarse, multiplicando su velocidad v por su tiempo de vida t, nos saldría que viajan en media unos 650 m. En otras palabras, se desintegran mucho antes de llegar a la superficie de la Tierra. Sin embargo en la práctica medimos una gran cantidad de muones a nivel del mar. La explicación es que por su velocidad relativista, su "reloj interno" corre más lento debido a la dilatación del tiempo. Al 99 % de la velocidad de la luz, el factor de corrección vale $\gamma \approx 9$, de modo que en realidad viven unas 9 veces más que un muón en reposo y una cantidad considerable de ellos sí llega a la superficie.

La historia es un poco diferente desde el punto de vista de los propios muones. Para ellos su reloj corre al ritmo normal (ya que están en reposo con respecto a sí mismos) y sólo tienen una vida media de 2,2 $\cdot\, 10^{-6}s$. ¿Cómo podemos compaginar esto con el hecho de que gran parte de ellos lleguen a la superficie? Para ellos, la Tierra (y por lo tanto la atmósfera) se mueve con una velocidad $v \sim 0,99\, c$ hacia ellos. Debido a este movimiento, la atmósfera sufre una contracción de Lorentz y parece 9 veces más delgada, unos 1700 m, suficiente para que una cantidad considerable llegue al suelo.

Los efectos debidos a la dilatación del tiempo según un observador, son debidos a la contracción de Lorentz según otro. Esto indica una conexión íntima entre el espacio y el tiempo, como veremos en las siguientes secciones. Un ejemplo de esta conexión es la simultaneidad de eventos en dos puntos distintos del espacio.

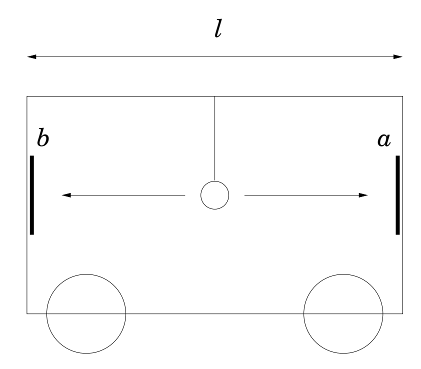


Figura 3.2: *La simultaneidad de eventos: una fuente de luz emite señales hacia dos detectores* a *y* b*, colocados al principio y al final del tren. Para* $\mathcal{O}'$ *las dos señales llegan simultáneamente a los detectores, mientras según* $\mathcal{O}$ *la señal llega primero al detector* b *y despues al detector* a.

Una fuente de luz, colocada en el medio del tren, emite dos señales hacia dos detectores a y b al principio y al final del tren respectivamente (Véase Figura 3.2). Para el pasajero en el tren $\mathcal{O}'$ las señales llegarán al mismo momento a los detectores, puesto que cada una de las señales tarda un tiempo $\Delta t' = l/2c$, donde l es la longitud del tren en reposo.

El observador $\mathcal{O}$ en el andén, sin embargo, ve también moverse las señales con la velocidad c, pero además ve que el detector b se aproxima con una velocidad v a la señal, mientras el detector a se aleja con velocidad v. Para él las señales tardan un tiempo

$$\Delta t_a = \frac{l\sqrt{1 - v^2/c^2}}{2(c - v)} \qquad \text{y} \qquad \Delta t_b = \frac{l\sqrt{1 - v^2/c^2}}{2(c + v)} \tag{3.6}$$

en llegar a los detectores a y b respectivamente. Para $\mathcal{O}$, la llegada de la señal al detector a no es simultánea a la llegada de la señal al detector b, sino que ocurre después de un intervalo de tiempo

$$\Delta t_a - \Delta t_b = \frac{lv/c^2}{\sqrt{1 - v^2/c^2}}. \tag{3.7}$$

Sucesos que son simultáneos para un observador, no necesariamente lo son para otro. Incluso es posible encontrar observadores $\mathcal{O}''$ para los cuales primero llega la señal en a y después en b, es decir, para los cuales el orden de los sucesos es invertido con respecto a $\mathcal{O}$: un observador que se mueve en la misma dirección que el tren, pero con velocidad $V > v$ con respecto al andén ve el tren moverse hacia atrás y por lo tanto el detector a acercarse a la señal, mientras b se aleja.

El hecho de que podamos encontrar distintos observadores que ven un suceso A antes, a la vez, o después de un suceso B, ¿puede causar problemas de causalidad? No, la relatividad es estrictamente causal. Sólo es posible invertir el orden de sucesos que no están en contacto causal, es decir para los cuales no da tiempo a mandar una señal de luz entre uno y otro. Esto se hará más claro cuando introduzcamos el concepto de conos de luz en la sección 3.5.

3.2. Las transformaciones de Lorentz

Nótese que todos los efectos relativistas derivados anteriormente (la dilatación del tiempo, la contracción de Lorentz y la no-simultaneidad de sucesos) son debidos al segundo postulado: la constancia de la velocidad de la luz. El primer postulado, la equivalencia de los observadores inerciales nos dirá cómo relacionar las mediciones de un observador con las de otro.

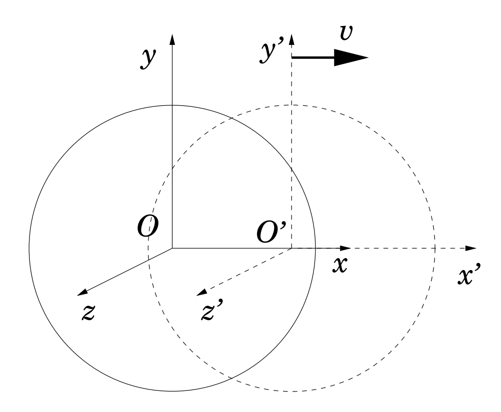


Figura 3.3: *Dos sistemas de referencia $\mathcal{O}$ y $\mathcal{O}'$ se mueven con una velocidad relativa v. Los sistemas de referencia están orientados tal que los ejes son paralelos y los relojes están sincronizados tal que en $t = t' = 0$ los orígenes coinciden. Debido a la constancia de la velocidad de la luz, una señal luminosa, emitida en todas las direcciones desde el origen en el momento $t = t' = 0$, será esférica en ambos sistemas de referencia.*

Un *evento* es algo que ocurre en cierto lugar del espacio en cierto momento. Un observador $\mathcal{O}$ medirá este suceso en su sistema de referencia y puede asignarle 4 coordenadas (x, y, z, t). Otro observador $\mathcal{O}'$ verá el mismo suceso y le asignará las coordenadas (x', y', z', t') en su propio sistema de referencia. La pregunta es: ¿cómo los dos observadores se pueden poner de acuerdo sobre sus mediciones? En otras palabras, ¿cómo están relacionadas las coordenadas (x, y, z, t) y (x', y', z', t'), si $\mathcal{O}'$ se mueve de manera uniforme y rectilínea con velocidad $\vec{v}$ con respecto a $\mathcal{O}$? Por simplicidad supondremos que los sistemas de referencia de $\mathcal{O}$ y $\mathcal{O}'$ están orientados de modo que los ejes son paralelos y que $\mathcal{O}'$ se mueve a lo largo del eje x de $\mathcal{O}$. Además, supondremos que en $t = t' = 0$ los dos orígenes de los sistemas de referencia coinciden (Véase Fig 3.3).

La mecánica newtoniana (y nuestra intuición) dice que la relación entre los dos sistema de referencia viene dada por las transformaciones de Galilei (2.8)

$$x' = x - vt, \qquad y' = y, \qquad z' = z, \qquad t' = t. \tag{3.8}$$

Sin embargo, es fácil ver que estas transformaciones no respetan el segundo postulado de la constancia de la velocidad de la luz. Como consecuencia, las transformaciones de Galilei están en contradicción con los resultados derivados en la sección anterior, en particular con la fórmula (3.3) de la dilatación del tiempo. Necesitamos por lo tanto derivar unas transformaciones nuevas que respeten el segundo postulado y reproduzcan los resultados de la sección anterior.

La transformación lineal más general[1] entre (x, t) y (x', t') (suponiendo que $y' = y$ y $z' = z$, dado que la contracción de Lorentz sólo es longitudinal) es

$$x' = Ax + Bt, \qquad t' = Cx + Dt, \tag{3.9}$$

donde los coeficientes A, B, C y D son funciones de v y c. A base de consideraciones generales de simetría, podemos relacionar algunos de los coeficientes. El origen del sistema $\mathcal{O}$ está en reposo con respecto a este sistema y satisface por lo tanto en cada momento la condición $x = 0$. Las transformaciones (3.9) en este caso se reducen a

$$x' = Bt, \qquad t' = Dt, \tag{3.10}$$

[1]En su famoso artículo *"Sobre la electrodinámica de cuerpos en movimiento"* en que presenta su teoría de la relatividad especial, Einstein decía que *"está claro que las ecuaciones deben ser lineales debido a las propiedades de homogeneidad que le asignamos al espacio y al tiempo."* Probablemente lo que quería decir es que esperaba que las transformaciones fueran las mismas en todos los puntos del espacio y por lo tanto la matriz de transformación una matriz constante. Veremos a partir del Capítulo 6 lo que ocurre si en vez de transformaciones globales consideramos transformaciones locales.

y, derivando la primera ecuación con respecto a t', tenemos

$$-v = B/D, \tag{3.11}$$

donde $-v = dx'/dt'$ es la velocidad con la que $\mathcal{O}'$ ve moverse el origen de $\mathcal{O}$.

De manera similar, el origen de $\mathcal{O}'$ está en reposo con respecto a sí mismo, de modo que $x' = 0$ y

$$Ax + Bt = 0, \tag{3.12}$$

o, derivando con respecto a t,

$$Av + B = 0. \tag{3.13}$$

Comparando (3.11) y (3.13) vemos que $A = D$ y $B = -vA$, de modo que (3.9) se convierte en

$$x' = A(x - vt), \qquad\qquad t' = A\,(C/A\,x + t). \tag{3.14}$$

Para determinar A y C, consideramos el siguiente suceso: en el momento $t = t' = 0$ (cuando los orígenes de ambos sistemas de referencia satisfacen la condición $x = x' = 0$), uno de los observadores emite una señal de luz desde el origen, que se expande en una esfera (véase Figura 3.3). Después de un tiempo t un punto en la esfera tendrá para $\mathcal{O}$ unas coordenadas que satisfagan

$$x^2 + y^2 + z^2 = (ct)^2. \tag{3.15}$$

Para $\mathcal{O}'$, la luz también se expande en una esfera, dado que la velocidad de la luz es igual para los dos, según el segundo postulado, sólo que para $\mathcal{O}'$ las coordenadas satisfarán

$$x'^2 + y'^2 + z'^2 = (ct')^2. \tag{3.16}$$

Dado que $y = y'$ y $z = z'$, vemos que una condición de consistencia para las transformaciones (3.14) es que

$$(ct)^2 - x^2 = (ct')^2 - x'^2. \tag{3.17}$$

Sustituyendo (3.14), encontramos para los valores de los coeficientes

$$A = \frac{1}{\sqrt{1 - v^2/c^2}}, \qquad\qquad C = \frac{-v/c^2}{\sqrt{1 - v^2/c^2}}. \tag{3.18}$$

Vemos por lo tanto que las transformaciones (3.9) entre observadores $\mathcal{O}$ y $\mathcal{O}'$ que respetan los dos postulados de la relatividad son de la forma

$$t' = \frac{t - vx/c^2}{\sqrt{1 - v^2/c^2}}, \qquad\qquad x' = \frac{x - vt}{\sqrt{1 - v^2/c^2}}. \tag{3.19}$$

Las transformaciones (3.19) se llaman las *transformaciones de Lorentz* y se reducen para $v \ll c$ a las transformaciones de Galilei (3.8) (por esto la mecánica newtoniana es válida a velocidades cotidianas). Obsérvese que las transformaciones (3.19) fueron construidas de tal manera que dejaran la cantidad

$$s^2 = (ct)^2 - x^2 - y^2 - z^2 \tag{3.20}$$

invariante (ejerc.). Dejamos como ejercicio derivar que las transformaciones de Lorentz reproducen exactamente los efectos relativistas de la sección 3.1.

3.3. Dinámica relativista

Al derivar las transformaciones de Lorentz hemos tenido que asumir que el factor $\sqrt{1 - v^2/c^2}$ es real, es decir, que la velocidad relativa entre dos observadores siempre es menor que la velocidad de la luz, $v < c$; sino las transformaciones (3.19) no están bien definidas.[2] Más general,

[2]Esto elimina el problema del capítulo anterior, donde un observador que se mueve junto con una onda electromagnética vería un perfil de campo electromagnético no permitido por las ecuaciones de Maxwell Ahora vemos que tal observador no existe, ya que para eso debería viajar con la velocidad de la luz.

las transformaciones de Lorentz implican que la velocidad de una partícula en un sistema de referencia dado siempre tiene que ser menor que (o, en caso de partículas sin masa, igual a) la velocidad de la luz. En otras palabras, la velocidad de la luz aparece en relatividad especial como un límite superior a cualquier velocidad permitida, una limitación que en la mecánica newtoniana no aparece.[3] Habrá que modificar por lo tanto la mecánica newtoniana de modo que respete este límite de velocidad.

Una primera observación es que la regla de la suma de las velocidades $V = V' + v$ no puede ser cierta, ya que permite velocidades mayores que c.[4] De las transformaciones de Lorentz podemos derivar una regla nueva para la composición de las velocidades.

Una partícula tiene una velocidad V' con respecto al observador $\mathcal{O}'$, que a su vez se mueve con velocidad v con respecto a $\mathcal{O}$. Para calcular la velocidad V de la partícula con respecto a $\mathcal{O}$, utilizamos la inversa de las transformaciones de Lorentz

$$x = \gamma\,(x' + vt'), \qquad\qquad t = \gamma\,(t' + vx'/c^2), \tag{3.21}$$

donde γ es el factor relativista, dado por (3.4). La velocidad $V = dx/dt$ viene dada por

$$V \;=\; \gamma\Big(\frac{dx'}{dt} + v\frac{dt'}{dt}\Big) \;=\; \gamma\,(V' + v)\,\frac{dt'}{dt} \;=\; \frac{V' + v}{1 + \frac{vV'}{c^2}}\,, \tag{3.22}$$

donde en la segunda igualdad hemos utilizado la regla de la cadena

$$\frac{dx'}{dt} = \frac{dx'}{dt'}\frac{dt'}{dt} = V'\frac{dt'}{dt}. \tag{3.23}$$

Es fácil ver que con esta *regla relativista de la suma de velocidades* (3.22), las velocidades medidas por $\mathcal{O}$ nunca exceden la velocidad de la luz. Obsérvese que incluso en el caso donde una partícula (sin masa) se mueve con velocidad $V' = c$ con respecto al observador $\mathcal{O}'$, su velocidad con respecto al obervador $\mathcal{O}$ según la regla (3.22) es $V = c$, independientemente de la velocidad v relativa entre $\mathcal{O}$ y $\mathcal{O}'$. Esto nos dice que la nueva regla de composición de velocidades respeta la constancia de la velocidad de la luz, exigido por el segundo postulado de la relatividad especial.

La segunda ley de Newton, $\vec{F} = m\vec{a}$ tampoco puede ser válida, puesto que implicaría que una partícula sometida a una fuerza constante podría alcanzar velocidades arbitrariamente grandes. Habrá que modificar por lo tanto la segunda ley, de modo que no viole los postulados o las transformaciones de Lorentz.

Se puede reescribir la segunda ley de Newton, definiendo la fuerza $\vec{F}$ como la derivada del momento $\vec{p}$ con respecto al tiempo

$$\vec{F} = \frac{d\vec{p}}{dt}, \tag{3.24}$$

sólo que ahora el momento $\vec{p}$ no es simplemente $m_0\vec{v}$, con m_0 la masa de la partícula (en reposo),[5] sino

$$\vec{p} = \frac{m_0\vec{v}}{\sqrt{1 - v^2/c^2}}. \tag{3.25}$$

[3]En 2011, los científicos trabajando en el experimento de OPERA en Gran Sasso (Italia) anunciaron una anomalía en sus datos experimentales potencialmente revolucionaria: unos neutrinos viajando con velocidades (ligeramente) mayores que la velocidad de la luz. Midiendo la velocidad de un haz de neutrinos producidos en el CERN en Ginebra (Suiza) y enviados a Gran Sasso, 731 km más lejos, parecían encontrar que los neutrinos llegaban 60 nanosegundos antes que si hubieran viajado con la velocidad de a luz en el vacío. Aunque inicialmente afirmaron que el error experimental sobre esta medida era de 20 nanosegundos (por lo tanto desafiando uno de los principios claves de la relatividad especial), un análisis más exhaustivo reveló que, debido a un cable de fibra óptica mal colocado, la barra de error real era más grande, concretamente de 150 nanosegundos, volviendo los resultados experimentales compatibles con la relatividad especial.

[4]Nótese que al derivar la regla de la suma de las velocidades (2.10) hemos supuesto implícitamente que el tiempo era el mismo para los dos observadores, $t = t'$, cosa que ahora sabemos que no es verdad.

[5]A veces se encuentra en los libros divulgativos el concepto de masa relativista $m_r = m_0/\sqrt{1 - v^2/c^2}$, sugiriendo que la masa de un objeto depende de su velocidad. Esta notación ahora está considerada anticuada y suele llevar a confusiones y errores conceptuales, por lo que, como en la gran mayoría de los libros de texto modernos, sólo usaremos la masa invariante m_0.

El factor γ en esta definición del momento da correcciones relativistas importantes a la mecánica newtoniana. Nótese por ejemplo que de este modo una fuerza nunca podrá acelerar una masa a velocidades más grandes que c, puesto que si $v \sim c$, hace falta una fuerza cada vez más grande para conseguir un aumento cada vez más pequeño en la velocidad.

En analogía con la mecánica newtoniana, definimos la diferencia de la energía cinética E como el trabajo realizado por la fuerza $\vec{F}$ a lo largo de una curva C

$$E_2 - E_1 = \int_C \vec{F} \cdot d\vec{r} = \int_C \vec{F} \cdot \vec{v} dt. \tag{3.26}$$

Utilizando la versión relativista de la segunda ley (3.24), podemos escribir la energía cinética como

$$\Delta E = \int_1^2 \vec{v} \cdot \frac{d\vec{p}}{dt} dt = \int_1^2 \vec{v} \cdot \frac{d\vec{p}}{d\gamma} \frac{d\gamma}{dt} dt = m_0 c^2 \int_1^2 \frac{d\gamma}{dt} dt = m_0 c^2 \Delta\gamma \tag{3.27}$$

donde en la tercera igualdad hemos utilizado que

$$\vec{v} \cdot \frac{d\vec{p}}{d\gamma} = \vec{v} \cdot \frac{d}{d\gamma}\Big(m_0 \gamma \vec{v} \Big) = m_0 v^2 + m_0 \gamma v \frac{dv}{d\gamma} = m_0 v^2 + m_0 \frac{c^2}{\gamma^2} = m_0 c^2. \tag{3.28}$$

La ecuación (3.27) da una relación entre la energía cinética E, la masa m_0 y la velocidad v que difiere bastante de la expresión newtoniana:

$$E = \frac{m_0 c^2}{\sqrt{1 - v^2/c^2}}. \tag{3.29}$$

Sin embargo, a bajas velocidades $v \ll c$, se puede desarollar el factor relativista en un desarrollo de Taylor

$$E \approx m_0 c^2 + \frac{1}{2} m_0 v^2 + ... \tag{3.30}$$

de modo que en el segundo término reconocemos la expresión newtoniana para la energía cinética, mientras el primer término es la *energía de reposo* que tiene un objeto simplemente por tener masa. Los puntos suspensivos dan correcciones relativistas a la expresión newtoniana de la energía cinética.

El hecho de que la energía de una partícula dependa de la velocidad, sugiere una relación profunda entre la energía y el momento. Un observador $\mathcal{O}'$ que está en reposo respecto a la partícula verá una energía $E' = m_0 c^2$ y un momento $\vec{p}\,' = 0$, mientras que el observador $\mathcal{O}$ verá un momento no-nulo y una energía más grande. Resulta que E y $\vec{p}$ están relacionados a través de una transformación de Lorentz (suponiendo que los sistemas de $\mathcal{O}$ y $\mathcal{O}'$ están orientados como en la sección anterior y que la masa se mueve a lo largo del eje x de $\mathcal{O}$):

$$E' = \frac{E - v p_x}{\sqrt{1 - v^2/c^2}}, \qquad p'_x = \frac{p_x - vE/c^2}{\sqrt{1 - v^2/c^2}}. \tag{3.31}$$

Finalmente, podemos combinar E y $\vec{p}$ en una expresión. De (3.25) y (3.29) tenemos que

$$E^2 - p_x^2 c^2 - p_y^2 c^2 - p_z^2 c^2 = m_0^2 c^4. \tag{3.32}$$

Obsérvese que el lado derecho es una constante, de modo que la combinación a la izquierda tiene el mismo valor para todos los observadores. Observe la similitud con la fórmula (3.17) de la distancia recorrida por la luz en un intervalo de tiempo.

En contraste con la mecánica newtoniana, la teoría de Maxwell del electromagnetismo no necesita correcciones relativistas, puesto que las leyes de Maxwell ya son invariantes bajo las transformaciones de Lorentz.

Las leyes de Maxwell

$$\vec{\nabla} \cdot \vec{E} = \rho, \qquad \vec{\nabla} \cdot \vec{B} = 0, \tag{3.33}$$

$$\vec{\nabla} \times \vec{E} = -\frac{1}{c}\, \partial_t \vec{B}, \qquad \vec{\nabla} \times \vec{B} = \frac{1}{c}\Big(\vec{\jmath} + \partial_t \vec{E}\Big), \tag{3.34}$$

junto con la fuerza de Lorentz

$$\vec{F} = e\Big(\vec{E} + \frac{\vec{v} \times \vec{B}}{c}\Big) \tag{3.35}$$

describen las interacciones de campos eléctricos $\vec{E}$, magnéticos $\vec{B}$, densidades de cargas ρ y densidades de corrientes $\vec{\jmath}$. Una de sus consecuencias es la ley de conservación de la carga

$$\partial_t \rho + \vec{\nabla} \cdot \vec{\jmath} = 0. \tag{3.36}$$

Está claro que el Principio de la Relatividad relaciona ciertas magnitudes que aparecen en las leyes de Maxwell. Por ejemplo, una corriente no es más que carga en movimiento, $\vec{\jmath} = \rho\vec{v}$, de modo que lo que para un observador $\mathcal{O}$ parece una corriente, es una carga estática para el observador $\mathcal{O}'$ que viaja con la misma velocidad v que la carga. Cargas y corrientes están por lo tanto relacionadas a través de una transformación de Lorentz (suponiendo que los sistemas de $\mathcal{O}$ y $\mathcal{O}'$ están orientados como en la sección anterior y que la carga se mueve a lo largo del eje x),

$$\rho' = \frac{\rho - v j_x/c^2}{\sqrt{1 - v^2/c^2}}, \qquad j'_x = \frac{j_x - v\rho}{\sqrt{1 - v^2/c^2}}. \tag{3.37}$$

Obsérvese que la ley de conservación de carga (3.36) es invariante bajo las transformaciones de Lorentz (3.19) y (3.37) (ejerc.). La conservación de la carga es por lo tanto válida para todos los observadores, tal como esperaríamos de una propiedad tan importante.

Según la ley de Gauss (3.33a), una carga estática causa un campo eléctrico estático, mientras la ley de Ampère (3.34b) dice que una corriente causa un campo magnético. Por lo tanto al pasar la carga, $\mathcal{O}$ verá un campo magnético que está ausente para $\mathcal{O}'$. Esto es posible porque los campos $\vec{E}$ y $\vec{B}$ se transforman bajo una transformación de Lorentz como

$$E'_x = E_x, \qquad E'_y = \frac{E_y - vB_z/c}{\sqrt{1 - v^2/c^2}}, \qquad E'_z = \frac{E_z + vB_y/c}{\sqrt{1 - v^2/c^2}},$$

$$B'_x = B_x, \qquad B'_y = \frac{B_y + vE_z/c}{\sqrt{1 - v^2/c^2}}, \qquad B'_z = \frac{B_z - vE_y/c}{\sqrt{1 - v^2/c^2}}. \tag{3.38}$$

Es un ejercicio instructivo comprobar que las ecuaciones de Maxwell transforman de manera covariante bajo las transformaciones (3.19), (3.37) y (3.38).

Como hemos visto en la sección 1.4, las ecuaciones homogéneas de Maxwell (3.33b) y (3.34a) implican que existen unos potenciales electromagnéticos ϕ y $\vec{A}$, tales que

$$\vec{E} = -\vec{\nabla}\phi - \frac{1}{c}\partial_t \vec{A}, \qquad \vec{B} = \vec{\nabla} \times \vec{A}, \tag{3.39}$$

y que podemos escribir las ecuaciones inhomogéneas de Maxwell (3.33a) y (3.34b) como

$$\frac{1}{c^2}\partial_t^2 \phi - \nabla^2 \phi = \rho, \qquad \frac{1}{c^2}\partial_t^2 \vec{A} - \nabla^2 \vec{A} = \frac{1}{c}\vec{\jmath}, \tag{3.40}$$

si suponemos que ϕ y $\vec{A}$ satisfacen el gauge de Lorenz

$$\frac{1}{c}\, \partial_t \phi + \vec{\nabla} \cdot \vec{A} = 0. \tag{3.41}$$

No es de extrañar por lo tanto que también los potenciales se transformen bajo las transformaciones de Lorentz

$$\phi' = \frac{\phi - vA_x/c}{\sqrt{1-v^2/c^2}}, \qquad A'_x = \frac{A_x - v\phi/c}{\sqrt{1-v^2/c^2}}. \tag{3.42}$$

También es un ejercicio instructivo demostrar que la condición de Lorenz (3.41) es invariante[6] y que las ecuaciones (3.40) son covariantes bajo las transformaciones de Lorentz (3.19), (3.37) y (3.42) y que se recupera las transformaciones (3.38) desde (3.42) y (3.39).

3.4. La fuerza de Lorentz y el Principio de la Relatividad

Puede resultar sorprendente que en la expresión (3.35) aparezca explícitamente la velocidad $\vec{v}$ de la partícula sometida a la fuerza de Lorentz, sobre todo después del énfasis que hemos dado en el capítulo anterior al Principio de la Relatividad. Parece que, para que tenga sentido la expresión, habrá que especificar con respecto a qué sistema de referencia se mide esta velocidad, violando así por lo tanto un primer principio que creíamos fundamental.

Sin embargo, curiosamente la expresión (3.35) para la fuerza de Lorentz es válida en todos los sistemas inerciales, ya que no sólo cambia $\vec{v}$ al cambiar de sistema de referencia, sino también $\vec{E}$ y $\vec{B}$, como acabamos de ver. No nos sorprenderá por lo tanto que la combinación (3.35) transforme adecuadamente bajo una transformación de Lorentz.

Una bonita ilustración de esta propiedad podemos obtener si analizamos el conocido efecto de las fuerzas entre dos conductores paralelos desde dos puntos de vista diferentes: desde el sistema de referencia del laboratorio, en el cual una corriente corre por los conductores, y desde el sistema que está en reposo con respecto a los electrones y en el que se mueve el conductor. La fuerzas que actúan sobre el sistema son las mismas, pero la interpretación en cada sistema es diferente (que es, como hemos visto ya, la esencia de la teoría de la relatividad).

En 1820 Ampère demostró ante la Academia de Ciencias de París que dos conductores paralelos, por los que pasan corrientes, experimentan una fuerza atractiva si las corrientes en ambos conductores van en el mismo sentido y una fuerza repulsiva si tienen sentidos opuestos. Visto desde el sistema de referencia del experimentador, la situación es bastante sencilla: los conductores están hechos de iones de cobre fijos (cargados positivamente) y electrones móviles (cargados negativamente), cuyo movimiento (que supondremos uniforme y rectilíneo) causa las corrientes (véase Figura 3.4). Puesto que los conductores son eléctricamente neutros, las densidades de las cargas positivas y negativas tienen que ser las mismas (módulo el signo de la carga): $\rho_+ = -\rho_-$. Por lo tanto los conductores no generan un campo eléctrico, pero cada una de las corrientes genererá un campo magnético del tipo (1.32), que afecta a los electrones moviéndose en el otro conductor. Un electrón (con carga $-e < 0$), que se mueve con una velocidad $\vec{v}$ dentro de un conductor, nota una fuerza de magnetostática

$$\vec{F} \;=\; -e\,\frac{\vec{v}\times\vec{B}}{c}, \tag{3.43}$$

donde $\vec{B}$ es el campo magnético causado por la corriente I en el otro conductor. Es fácil de comprobar que efectivamente la fuerza es atractiva si $\vec{v}$ y $\vec{I}$ son antiparalelos (es decir, si las dos corrientes tienen el mismo sentido) y repulsiva si $\vec{v}$ y $\vec{I}$ son paralelos (si las corrientes van en

[6] El gauge de Lorenz recibe su nombre por Ludwig Lorenz (1829-1891), el físico danés que lo introdujo en 1867 y no por H.A. Lorentz, el físico holandés, conocido por la fuerza de Lorentz y el grupo de Lorentz. La jugada graciosa de la Historia es que el particular interés del gauge de Lorenz está en el hecho de que es invariante Lorentz.

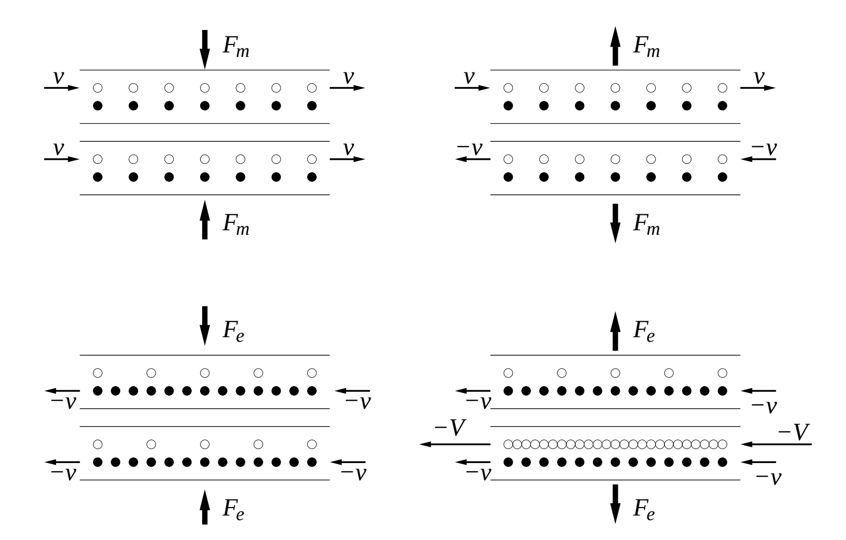


Figura 3.4: *Las fuerzas electrostáticas y magnetostáticas entre dos conductores paralelos vistos desde dos sistemas de referencia distintos: En el sistema del laboratorio (arriba) la densidad de cargas positivas (círculos negros) es igual (pero opuesta) a la densidad de cargas negativas (círculos blancos) y la única fuerza que experimentan los electrones es magnetostática, debido a que se mueven en el campo magnético del otro conductor. La fuerza es atractiva si las corrientes son paralelas (izquierda) y repulsiva si son antiparalelas (derecha). En el sistema de referencia de los electrones (abajo), debido a la contracción de Lorentz, el otro conductor tiene una densidad de carga neta positiva si las dos corrientes son paralelas (izquierda) y negativa si son antiparalelas (derecha), lo que resulta en unas fuerzas electrostáticas atractivas y repulsivas respectivamente.*

direcciones opuestas).[7]

Desde el sistema de referencia de los electrones se ve la situación distinta, además hay que distinguir entre el caso de las corrientes paralelas y antiparalelas. Si las dos corrientes van en la misma dirección, entonces desde el punto de vista del sistema de los electrones, los electrones de los dos conductores están en reposo, mientras que los iones se mueven con una velocidad $-v$. En este sistema de referencia, las distancias entre los iones están contraídas con un factor $\sqrt{1-v^2/c^2}$ por la contracción de Lorentz, de modo que la densidad de carga posiva ρ'_+ es más alta que en el sistema de laboratorio,[8]

$$\rho'_+ = \frac{\rho_+}{\sqrt{1-v^2/c^2}}. \tag{3.44}$$

Por otro lado, la densidad de carga negativa ρ'_- en este sistema es más baja que en el sistema de laboratorio, ya al estar en reposo ya no sufren la contracción de Lorentz que experimentaban en el sistema de laboratorio. Concretamente tenemos que

$$\rho'_- = \rho_- \sqrt{1-v^2/c^2}, \tag{3.45}$$

y por lo tanto la densidad de carga total de un conductor visto por un electrón en el otro es

$$\rho'_{\rm tot} \;=\; \rho'_+ \,+\, \rho'_- \;=\; \frac{\rho_+}{\sqrt{1-v^2/c^2}} \;+\; \rho_- \sqrt{1-v^2/c^2} \;=\; \frac{\rho_0\, v^2/c^2}{\sqrt{1-v^2/c^2}}, \tag{3.46}$$

[7]Escribiremos la corriente I en terminos de ρ_+. Es bien conocido el error histórico del científico y político estadounidense Benjamin Franklin (1706 - 1790) de identificar la corriente eléctrica como el movimiento de cargas positivas. El descubrimiento del electrón en 1897 por el físico inglés Joseph John Thomsom (1856 - 1940) aclaró el asunto, pero ya era tarde para cambiar una notación generalmente aceptada, de modo que sigue el convenio que $\vec{I} = -\rho_-\vec{v}A$, con A la superficie seccional del conductor.

[8]Nótese que esto es un caso especial de la fórmula (3.37), donde $(j_+)_x = 0$ en el sistema de laboratorio.

donde en la última igualdad hemos utilizado que en el sistema de laboratorio $\rho_+ = -\rho_- = \rho_0$. En otras palabras un electrón en un conductor verá que el otro conductor tiene una carga eléctrica neta positiva y que su campo eléctrico ejerce una fuerza atractiva sobre él

$$\vec{F} \;=\; -e\,\vec{E}. \tag{3.47}$$

Obviamente el electrón nota también el campo magnético de la corriente de iones en el otro conductor, pero ese campo magnético no ejerce ninguna fuerza sobre él, ya que está en reposo en su propio sistema de referencia.

Si las dos corrientes son antiparalelas, la situación es un poco más complicada: un electrón ve moverse a los iones en el otro conductor con una velocidad $-v$ y nota por lo tanto una densidad de carga positiva como en (3.44). Pero también ve, por la regla de sumatorio de velocidades (3.22), a los electrones moverse con una velocidad $V = 2v/(1+v^2/c^2)$ y por lo tanto las distancias entre éstos aún más Lorentz contraídos que los iones. La densidad de carga negativa es por lo tanto más alta que la positiva, y el otro conductor tiene una densidad de carga global negativa, lo que resulta en una fuerza electrostática repulsiva.

Por lo tanto vemos que la fuerza que actúa sobre un electrón en un conductor es puramente magnetostática desde el punto de vista del experimentador en el laboratorio, pero puramente electrostática en el sistema de referencia de los electrones, aunque ambas fuerzas están descritas por la misma fuerza de Lorentz. La misma expresión (3.35) es por lo tanto válida en todos los sistemas inerciales, independientemente del estado de movimiento de la partícula en cuestión.

Incluso se puede hacer este análisis de manera cuantitativa, aunque nos limitaremos a calcular la fuerza actuando sobre un solo electrón en un conductor por los campos eléctricos y magnéticos del otro conductor en el caso de las corrientes paralelas. Dejaremos el otro caso como ejercicio para el estudiante interesado.

Como hemos dicho, el campo magnético de un conductor lineal por el que pasa una corriente $\vec{I} = \pi\rho_0 v\, R_0^2\, \vec{e}_z$, con R_0 el radio del conductor, viene dado por (1.32),

$$\vec{B} \;=\; \frac{\rho_0 v R_0^2}{2cr}\,\vec{e}_\varphi, \tag{3.48}$$

y por lo tanto un electrón que se mueve en el sistema del laboratorio con una velocidad $\vec{v} = -v\vec{e}_z$ antiparalela a la corriente $\vec{I}$ (es decir, forma parte de una corriente que va en el mismo sentido que I), experimenta una fuerza de Lorentz

$$\vec{F} \;=\; -e\,\frac{\vec{v}\times\vec{B}}{c} \;=\; \frac{-e\rho_0 v^2 R_0^2}{2c^2 r}\,\vec{e}_r. \tag{3.49}$$

Por otro lado, en el sistema de referencia del electrón, un conductor lineal con densidad de carga $\rho'_{\rm tot}$ crea un campo eléctrico (ejerc.)

$$\vec{E} \;=\; \frac{R_0^2\,\rho'_{\rm tot}}{2r}\vec{e}_r, \tag{3.50}$$

por lo que el electrón experimenta una fuerza de Lorentz

$$\vec{F}' \;=\; -e\vec{E} \;=\; \frac{1}{\sqrt{1-v^2/c^2}}\,\frac{-e\rho_0 v^2 R_0^2}{2c^2 r}\,\vec{e}_r, \tag{3.51}$$

donde hemos usado la fórmula (3.46) que relaciona la densidad $\rho'_{\rm tot}$ con ρ_0. Vemos por lo tanto que las expresiones para la fuerza de Lorentz en ambos sistemas de referencia están relacionadas como

$$F' \;=\; \frac{1}{\sqrt{1-v^2/c^2}}\,F. \tag{3.52}$$

Esta relación obviamente no es casualidad: esta fórmula claramente es un caso especial de una transformación de Lorentz entre la fuerza y otra cantidad, que en este caso es cero en el sistema del laboratorio. Efectivamente, en el Capítulo 5 veremos que la fuerza y el trabajo W ejercido por esa fuerza por unidad de tiempo transforman bajo una transformación de Lorentz como

$$W' = \frac{W - vF_x/c^2}{\sqrt{1 - v^2/c^2}}, \qquad\qquad F'_x = \frac{F_x - vW}{\sqrt{1 - v^2/c^2}}. \tag{3.53}$$

En particular, como la fuerza de Lorentz en el sistema del laboratorio es puramente magnetostática, el trabajo ejercido por ella es cero en este sistema de referencia.

3.5. El espacio de Minkowski

Despúes de haber estudiado las reglas de transformación de diversas cantidades físicas, tanto cinemáticas como electrodinámicas, nos debería la llamar la atención que ciertas combinaciones matemáticas aparezcan en varias ocasiones. En realidad esto nos indica la existencia de una estructura matemática subyacente, que nos permitirá no sólo simplificar las ecuaciones, sino también nos enseñará una importante lección sobre la estructura del espaciotiempo. Estudiaremos en mucho detalle esta estructura matemática en el Capítulo 5, pero queremos terminar este capítulo con un pequeño adelanto.

Hemos visto en la sección 3.2 que una condición clave en la derivación de las transformaciones de Lorentz era el hecho de que éstas dejan la cantidad

$$s^2 = c^2t^2 - x^2 - y^2 - z^2 \tag{3.54}$$

invariante (véase por ejemplo (3.17)). En otras palabras, s^2 tiene el mismo valor para todos los observadores relacionados a través de una transformación de Lorentz. Por la forma del invariante s^2, podemos distinguir 3 casos: dependiendo de los valores de las coordenadas, s^2 puede ser mayor, menor o igual a cero.

Podemos representar todos los eventos en el *espacio de Minkowski*, un espacio de sucesos, donde cada suceso está caracterizado por su posición x, y, z y el momento t en que ocurre. En este espacio de sucesos, la ecuación $s^2 = 0$ es la trayectoria de una señal de luz esférica emitida desde el origen en el instante $t = 0$: en cada momento la señal consiste en una esfera de radio ct. Si dibujamos las coordenadas $x(t)$, $y(t)$, $z(t)$ como función del tiempo, vemos que la trayectoria corresponde a un cono invertido con el vértice en el origen $ct = x = y = z = 0$ (véase Figura 3.5). De igual modo la trayectoria de toda la luz que llega en el momento $t = 0$ al punto $x = y = z = 0$ es otro cono con el vértice en el origen. A estos dos conos se les llama el *cono de luz futuro* (zona I en Figura 3.5) y el *cono de luz pasado* (zona II en Figura 3.5) respectivamente.

La trayectoria de un observador $\mathcal{O}$ en el origen de su sistema de referencia, que utiliza las coordenadas (ct, x, y, z) es $x = y = z = 0$, es decir el eje ct, dado que está en reposo con respecto a sí mismo. La trayectoria de un observador $\mathcal{O}'$ que se mueve con velocidad constante con respecto a $\mathcal{O}$ y que coincide con $\mathcal{O}$ en el momento $t = 0$, es una recta que pasa por el origen (véase Figura 3.5). Debido a que la velocidad de $\mathcal{O}'$ siempre es menor que c, su trayectoria cae dentro del cono de luz. En general, debido a que c es la velocidad máxima permitida, cualquier partícula que pase por el origen siempre se quedará dentro del cono de luz.

Los puntos dentro del cono de luz satisfacen la condición $s^2 > 0$, mientras que los puntos fuera del cono de luz (la zona III en Figura 3.5) satisfacen la condición $s^2 < 0$. Cualquier punto dentro del cono de luz está en contacto causal con el punto $ct = x = y = z = 0$, puesto que es posible mandar una señal desde un punto dentro del cono pasado al origen, o desde el origen a un punto dentro del cono futuro. Sin embargo los puntos fuera del cono de luz no están en contacto causal con el origen: ninguna señal de luz llegaría desde el origen a un punto en la

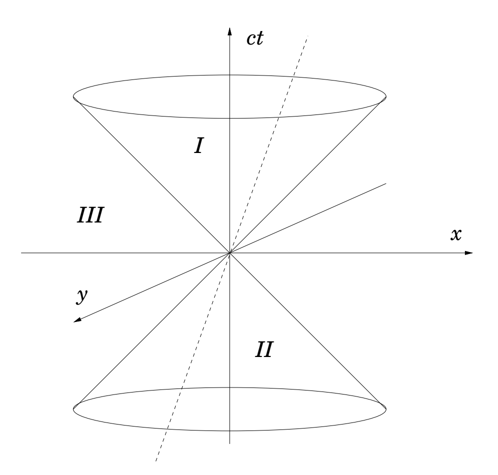


Figura 3.5: *Los conos de luz en el espacio de Minkowski: los conos representan la trayectoria de los rayos de luz que pasan por el punto $x = y = z = 0$ en el momento $t = 0$. El interior de los conos de luz (las zonas I y II) representan posibles trayectorias de observadores inerciales que se mueven con velocidad constante pasando por el origen (linea discontinua). El exterior de los conos (la zona III) no está en conexión causal con el origen.*

zona III o vice versa. Esta estructura causal es igual para todos los observadores, dado que las transformaciones de Lorentz dejan la cantidad s^2 invariante.

Matemáticamente el espacio de Minkowski tiene la esctructura de un espacio vectorial, de modo que podemos pensar en las coordenadas (ct, x, y, z) de un suceso como un vector cuadrimensional $\hat{x}$ en este espacio vectorial y en la cantidad s^2 como el cuadrado de la norma del vector. Fíjese que para esto hemos tenido que introducir una nueva definición para el producto escalar entre dos vectores $\hat{a}$ y $\hat{b}$ en el espacio de Minkowski como

$$\hat{a} \cdot \hat{b} = a_t b_t - a_x b_x - a_y b_y - a_z b_z. \tag{3.55}$$

Esta definición proporciona un producto escalar y una norma que no es definida positiva. El espacio de Minkowski por lo tanto no tiene una geometría euclídea, sino lo que se llama *pseudo-euclídea* o *lorentziana*. Volveremos a esto en más detalle en el Capítulo 5.

Por lo tanto la cantidad $s^2 = c^2t^2 - x^2 - y^2 - z^2$ representa el cuadrado de la distancia entre un suceso (ct, x, y, z) y el origen. En general, la cantidad

$$\Delta s^2 = c(t_2 - t_1)^2 - (x_2 - x_1)^2 - (y_2 - y_1)^2 - (z_2 - z_1)^2 \tag{3.56}$$

es el cuadrado de la distancia entre dos sucesos (ct_1, x_1, y_1, z_1) y (ct_2, x_2, y_2, z_2) en el espacio de Minkowski. Si la distancia $\Delta s^2 > 0$, los sucesos están separados por un *intervalo temporal*, si la distancia $\Delta s^2 = 0$ por un *intervalo nulo* o *tipo luz* y si la distancia $\Delta s^2 < 0$ por un *intervalo espacial*. Como hemos visto antes, sólo hay conexión causal entre sucesos separados por intervalos temporales o nulos.

Una transformación de Lorentz (3.19) relaciona las componentes (ct, x, y, z) de un vector de posición visto por un observador $\mathcal{O}$ con las componentes (ct', x', y', z') del mismo vector visto por otro observador $\mathcal{O}'$. Una transformación de Lorentz entonces no es más que un cambio de base dentro del espacio de Minkowski. Ya hemos visto que la trayectoria de $\mathcal{O}'$ es una recta por el origen dentro del cono de luz, donde el ángulo β entre la trayectoria y el eje ct es una medida de la velocidad de $\mathcal{O}'$: $\operatorname{tg} \beta = v/c$ (véase Figura 3.6). El observador $\mathcal{O}'$ tomará esta recta como su eje tempotal ct', puesto que está en resposo con respecto a sí mismo. Dado que para $\mathcal{O}'$ la velocidad de la luz tiene que ser igual que para $\mathcal{O}$, el eje x' de $\mathcal{O}'$ está orientado de manera simétrica con

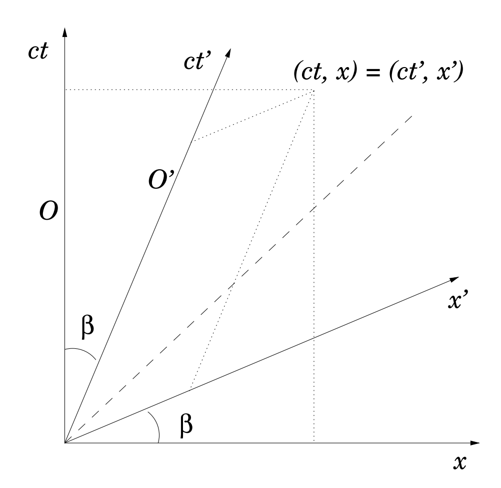


Figura 3.6: *Una transformación de Lorentz en el espacio de Minkowski: el mismo suceso tiene coordenadas* (ct, x) *para el observador* $\mathcal{O}$ *y coordenadas* (ct', x') *para el observador* $\mathcal{O}'$*. Los dos sistemas de referencia están relacionados a través de una transformación de Lorentz (3.19).*

respecto al cono de luz, es decir el ángulo entre el eje x y x' también es β (véase Figura 3.6). Tanto el sistema de referencia (ct, x, y, z) de $\mathcal{O}$ como el sistema (ct', x', y', z') de $\mathcal{O}'$ forman una base completa del espacio de Minkowski, y la diferencia entre ellos es un cambio de base a través de la transformación (3.19).[9]

Con los diagramas de espacio-tiempo descritos arriba podemos fácilmente derivar efectos como la contracción de Lorentz y la no-simultaneidad de eventos. Una varilla de longitud L' (medida en coordenadas x') que se mueve junto con el observador $\mathcal{O}'$, barrerá una trayectoria como la dibujada en la Figura 3.7. Del dibujo está claro que $\mathcal{O}$ y $\mathcal{O}'$ no ven la vara de la misma manera.

Para cada observador dos eventos simultáneos son dos eventos que tienen el mismo valor de la coordenada temporal. Pero como $\mathcal{O}$ y $\mathcal{O}'$ usan cada uno su propio coordenada temporal, no estarán de acuerdo sobre si ciertos eventos coinciden o no. Para $\mathcal{O}$, los sucesos simultáneos son los que tienen el mismo valor para ct (líneas horizontales en la Figura 3.7), como los puntos a y c a los extremos de la varilla visto por $\mathcal{O}$. Para $\mathcal{O}'$, los sucesos simultáneos tienen el mismo valor de ct' (líneas inclinadas en la Figura 3.7), como los puntos a y b a los extremos de la varilla visto por $\mathcal{O}'$. En otras palabras, para $\mathcal{O}$ en un momento dado la vara se extiende de a a c, mientras para $\mathcal{O}'$ se extiende de a a b. La vara vista por $\mathcal{O}'$ en un tiempo dado es un "corte de la vara visto por $\mathcal{O}$ en distintos momentos", y vice versa. Por lo tanto no es de extrañar que los observadores tampoco se ponen de acuerdo sobre la longitud de la vara: $\mathcal{O}$ mide una longitud L (el intervalo $[a, c]$), mientras que $\mathcal{O}'$ sólo mide L' (el intervalo $[a, b]$). La contracción de Lorentz y la no-simultaneidad de eventos van mano en mano.

En general, en el diagrama de la Figura 3.7 podemos ver que para cualesquiera dos eventos separados por un intervalo temporal existe un observador que ve estos eventos en el mismo lugar, mientras para dos eventos separados por un intervalo espacial existe un observador que los ve simultáneos (ejerc.).

Hemos visto que la posición (ct, x, y, z) de un evento es un vector en el espacio de Minkowski y la transformación de Lorentz (3.19), un cambio de base. Pero t y x no son las únicas cantidades físicas que están relacionadas a través de una transformación de Lorentz. La ecuación (3.31) dice que también la energía E y el momento $\vec{p}$ definido en (3.25) transforman unos en

[9] Obsérvese que, al contrario de lo que podría sugerir la Figura 3.6, tanto el sistema (ct, x, y, z) de $\mathcal{O}$ como el sistema (ct', x', y', z') de $\mathcal{O}'$ forman bases ortogonales según la definición (3.55) (ejerc.).

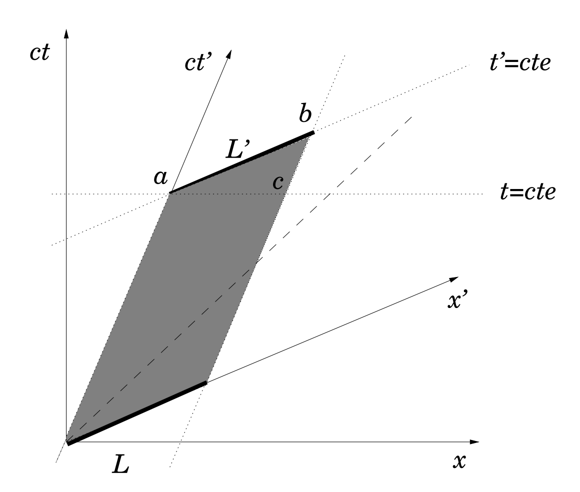


Figura 3.7: *La contracción de Lorentz y la no-simultaneidad de eventos en el espacio de Minkowski. Una varilla de longitud L' para un observador en reposo con respecto a la varilla, tiene una longitud $L < L'$ para un observador que ve la varilla moverse. Los eventos que un observador llama simultáneos no lo son para otro.*

otros. En otras palabras, E y $\vec{p}$ también forman las componentes de un vector cuadrimensional $\hat{p} = (E/c, p_x, p_y, p_z)$ en el espacio de Minkowski y transforman como (3.31) bajo el cambio de base (3.19). La relación (3.32) entonces es precisamente el cuadrado de la norma del vector de energía-momento $\hat{p} \cdot \hat{p} = m_0^2 c^2$, según la definición (3.55) del producto escalar. El hecho de que $E^2 - p^2c^2$ tenga el mismo valor para todos los observadores es justo porque es la norma de un vector cuadrimensional, y por lo tanto es, igual que s^2, un invariante Lorentz.

También en electromagnetismo hemos encontrado varias cantidades que transforman bajo transformaciones de Lorentz y que por lo tanto forman vectores cuadrimensionales: la carga y la corriente son componentes del vector $\hat{\jmath} = (c\rho, j_x, j_y, j_z)$ y de igual manera ϕ y $\vec{A}$ combinan para formar el vector $\hat{A} = (\phi, A_x, A_y, A_z)$. El caso de los campos electromagnéticos $\vec{E}$ y $\vec{B}$ es un poco más sutil, puesto que las transformaciones (3.38) son más complicadas que las transformaciones de Lorentz que hemos visto en otros casos. Resulta que $\vec{E}$ y $\vec{B}$ no son las componentes espaciales de dos vectores cuadrimensionales, sino que combinan en un tensor antisimétrico

$$\hat{F} = \begin{pmatrix} 0 & -E_x & -E_y & -E_z \\ E_x & 0 & -B_z & B_y \\ E_y & B_z & 0 & -B_x \\ E_z & -B_y & B_x & 0 \end{pmatrix}. \tag{3.57}$$

Hemos visto por lo tanto que toda la dinámica relativista, tanto la mecánica como el electromagnetismo, se puede formular en función de vectores y tensores cuadrimensionales que transforman de determinada manera bajo las transformaciones de Lorentz. Esto es necesario para que se cumpla el Principio de la Relatividad, es decir que las leyes de la física tengan la misma forma para todos los observadores en movimiento uniforme y rectilíneo relativo.

En el Capítulo 5 volveremos a introducir estos conceptos de manera más formal, que tiene la ventaja de ser directamente generalizable a transformaciones que relacionan observadores en movimiento no necesariamente uniforme y rectilíneo. Pero primero repasaremos un poco de álgebra lineal, necesario para entender bien este formalismo.

Capítulo 4

Álgebra de tensores y transformaciones ortogonales

La educación matemática del joven físico [Albert Einstein] *no era muy sólida, y yo estoy en una buena posición para evaluarla, puesto que la ha recibido de mí en Zürich hace tiempo.*
(H. Minkowski)

En este capítulo repasaremos las propiedades básicas de álgebra lineal y transformaciones ortogonales, ya que la formulación covariante de la relatividad especial recurre a las mismas técnicas.

4.1. El convenio de sumación de Einstein

A partir de este capítulo utilizaremos el llamado *convenio de sumación de Einstein*, una manera compacta de anotar fórmulas, introducida por Einstein en 1916. En cuanto un índice aparezca repetido arriba y abajo, se supone una suma sobre todos los posibles valores de este índice, suprimiendo un signo de sumatorio explícito. Los índices repetidos también se llaman *mudos*. La descomposición de un vector $|a\rangle$ en una base $\{|e_i\rangle\}$ sería por lo tanto $|a\rangle = a^i|e_i\rangle$ y un producto escalar entre dos vectores $\langle a|b\rangle = a_i b^i$, donde la suma en ambos casos contiene N términos en un espacio vectorial N-dimensional. Del mismo modo el producto de dos matrices $C = AB$ se anota entonces como $C^i{}_j = A^i{}_k B^k{}_j$.

Dado que un índice repetido sólo es un índice de sumación, da igual el nombre que le demos, de modo que tenemos que $a_i b^i = a_k b^k$.[1] Además, como dentro de cada término de la suma, los factores son simplemente números reales o complejos, el órden de los factores no influye, siempre y cuando respectemos el lugar de los índices: $C^i{}_j = A^i{}_k B^k{}_j = B^k{}_j A^i{}_k$, pero $A^i{}_k B^k{}_j \neq B^i{}_k A^k{}_j$.

Los índices que no están sumados se llaman *libres*. En una ecuación vectorial o tensorial, los índices libres en cada lado de la igualdad tienen que ser los mismos. El convenio de sumación de Einstein también supone N ecuaciones para cada índice libre (una para cada valor del índice).

[1] Siempre y cuando no se repita el nombre de índices contraídos: no es posible saber si la expresión errónea $a_i b_i c^i d^i$ se refiere a $\langle a|c\rangle\langle b|d\rangle = a_i b_j c^i d^j$ o a $\langle a|d\rangle\langle b|c\rangle = a_i b_j c^j d^i$.

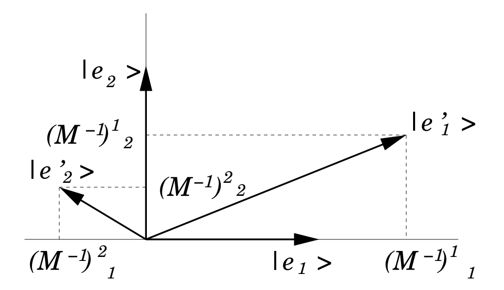


Figura 4.1: *La descomposición de los vectores de base $|e_j{}'\rangle$ en la base $\{|e_i\rangle\}$. Las componentes $(M^{-1})^i{}_j$ forman una matriz $N \times N$ que parametriza el cambio de base.*

4.2. Espacios vectoriales y duales

Consideramos el espacio vectorial N-dimensional $\mathbb{R}^N$, con un origen O y una base $\{|e_i\rangle\}$, formado por N vectores linealmente independientes. A cada punto x de $\mathbb{R}^N$ se le asigna un vector de posición $|x\rangle$, que se descompone en la base $\{|e_i\rangle\}$ como

$$|x\rangle \ = \ x^1|e_1\rangle + x^2|e_2\rangle + ... + x^N|e_N\rangle \ = \ x^i|e_i\rangle. \tag{4.1}$$

(Nótese que estamos utilizando el convenio de sumación.) La descomposición $|x\rangle$ en la base $\{|e_i\rangle\}$ es única y los números x^i son las *componentes de* $|x\rangle$ *en la base* $\{|e_i\rangle\}$. Otra manera de representar el vector $|x\rangle$ es a través de sus componentes, como un vector de columna

$$|x\rangle \ = \ \begin{pmatrix} x^1 \\ x^2 \\ \vdots \\ x^N \end{pmatrix}. \tag{4.2}$$

En vez de trabajar en la base $\{|e_i\rangle\}$, podríamos haber escrito $|x\rangle$ en una base distinta $\{|e_i{}'\rangle\}$, donde la descomposición sería

$$|x\rangle \ = \ x'^i|e_i{}'\rangle. \tag{4.3}$$

Para encontrar la relación entre las dos bases, descomponemos los vectores de base $|e_i{}'\rangle$ en la base $\{|e_i\rangle\}$ (véase Figura 4.1),

$$|e_j{}'\rangle = (M^{-1})^i{}_j \, |e_i\rangle, \tag{4.4}$$

donde $(M^{-1})^i{}_j$ es la componente i del vector $|e_j{}'\rangle$ en la base $\{|e_i\rangle\}$.[2] Desde el punto de vista matemático M^{-1} es la matriz $N \times N$ que parametriza la transformación entre las dos bases.

Dado que el vector $|x\rangle$ en (4.1) y (4.3) es el mismo y que la relación entre las dos bases viene dada por (4.4), tiene que haber una relación entre las componentes x^i y x'^i. No es difícil ver que las componentes transforman como

$$x'^i = M^i{}_j \, x^j, \tag{4.5}$$

donde la matriz M es la inversa de M^{-1},

$$M^i{}_j(M^{-1})^j{}_k \ = \ \delta^i{}_k \ = \ (M^{-1})^i{}_j M^j{}_k, \tag{4.6}$$

donde $\delta^i{}_j$ son las componentes de la matriz identidad $\mathbb{1}$

$$(\mathbb{1})^i{}_j = \delta^i{}_j = \begin{cases} 0 & \text{cuando } \ i \neq j, \\ 1 & \text{cuando } \ i = j. \end{cases} \tag{4.7}$$

[2] Aquí estamos suponiendo por simplicidad que las dos bases tienen el mismo origen. Si no es así, la transformación es de la forma $|e_j{}'\rangle = (M^{-1})^i{}_j \, |e_i\rangle + |a\rangle$.

Más que elementos de un espacio vectorial, para un físico un vector es un objeto con ciertas reglas de transformación. Por eso, en la física se suele tomar la regla de transformación como la definición de un vector: Cualquier objeto con N componentes x^i que bajo un cambio de base (4.4) transforma como (4.5) se le llama un *vector de columna* o un *vector contravariante*. (Hemos añadido los adjetivos "de columna" y "contravariante" para distinguirlo de otro tipo de vector que introduciremos en seguida.)

El hecho de que en la transformación de los vectores de base (4.4) aparezca la matriz M^{-1}, mientras en la transformación de las componentes del vector aparezca la inversa M, puede ser un poco confuso, pero tiene que ver con la diferencia entre una transformación activa y una pasiva. No entraremos en detalles aquí, puesto que no nos hace falta. En realidad toda la información física de un vector está contenida en sus componentes, de modo que dadas las componentes de un vector en una base y sabiendo la regla (4.5) de transformación a otra base, podemos trabajar sólo con las componentes, sin tener que hacer referencia a los vectores de base o las transformaciones de estos últimos. Esto es lo que haremos en este curso, salvo en esta sección y la siguiente, por razones pedagógicas.

Considera ahora el espacio de las aplicaciones lineales $\langle y|$ que llevan los vectores de $\mathbb{R}^N$ a los números reales $\mathbb{R}$: la aplicación $\langle y|$ actúa sobre el vector contravariante $|x\rangle$ como $\langle y|x\rangle \in \mathbb{R}$. Dado que estas aplicaciones son lineales (por construcción) tenemos que

$$\langle y|\Big(\alpha|x_1\rangle + \beta|x_2\rangle\Big) \;=\; \alpha\langle y|x_1\rangle + \beta\langle y|x_2\rangle. \tag{4.8}$$

Además una combinación lineal de dos de estas aplicaciones, también es una aplicación

$$\Big(\alpha\langle y_1| + \beta\langle y_2|\Big)|x\rangle \;=\; \alpha\langle y_1|x\rangle + \beta\langle y_2|x\rangle, \tag{4.9}$$

de modo que el espacio de las aplicaciones lineales también tiene la estructura matemática de un espacio vectorial, usualmente llamado el *espacio dual* ${}^*\mathbb{R}^N$. Por lo tanto podemos considerar también los $\langle y|$ como un tipo de vectores (aunque distinto de los $|x\rangle$, como veremos enseguida) y construir una base dual $\{\langle e^i|\}$, en la cual las aplicaciones lineas se descomponen como

$$\langle y| \;=\; y_1\langle e^1| + y_2\langle e^2| + ... + y_N\langle e^N| \;=\; y_i\langle e^i|. \tag{4.10}$$

Dada la estructura de espacio vectorial dual, es habitual llamar a los $\langle y|$ *vectores covariantes*, o *uno-formas*. Una manera alternativa de representar los vectores covariantes es por lo tanto como vector de fila:

$$(\, y_1,\ y_2,\ ...\ ,\ y_N) \tag{4.11}$$

Obsérvese que anotamos las componentes de los vectores contravariantes (es decir elementos de $\mathbb{R}^N$) con índice arriba, mientras las componentes de los vectores covariantes (los elementos del espacio dual) con índice abajo.

Los vectores covariantes actúan sobre los contravariantes mediante el *producto escalar*. Lo más sencillo es definir el producto escalar utilizando los vectores de base: definimos el producto escalar entre el vector de base dual $\langle e^i|$ y el vector de base $|e_k\rangle$ como

$$\langle e^i|e_j\rangle = \delta^i{}_j, \tag{4.12}$$

donde $\delta^i{}_j$ es la *delta de Kronecker*

$$\delta^i{}_j = \begin{cases} 0 & \text{cuando } \; i \neq j, \\ 1 & \text{cuando } \; i = j. \end{cases} \tag{4.13}$$

Utilizando las propiedades de linealidad (4.8) y (4.9), es obvio que en general el producto escalar entre un vector covariante $\langle y|$ y un vector contravariante $|x\rangle$ viene dado por

$$\langle y|x\rangle \;=\; y_i x^k\langle e^i|e_k\rangle \;=\; y_i x^k\delta^i{}_k \;=\; y_i x^i, \tag{4.14}$$

lo que efectivamente es un elemento de $\mathbb{R}$.

Lo que nos interesa en este momento es saber cómo cambian las componentes y_i del vector covariante bajo el cambio de coordenadas (4.4). Dado que por la definición del producto escalar (4.12) las componentes y_i en la base $\{|e_i\rangle\}$ vienen dadas por $y_i = \langle y|e_i\rangle$, tenemos que las componentes y'_i en la base $\{|e'_i\rangle\}$ toman la forma

$$y'_j \;=\; \langle y|e'_j\rangle \;=\; \langle y|(M^{-1})^i{}_j|e_i\rangle \;=\; (M^{-1})^i{}_j\langle y|e_i\rangle \;=\; (M^{-1})^i{}_j y_i, \tag{4.15}$$

o, resumiendo

$$y'_j \;=\; (M^{-1})^i{}_j y_i. \tag{4.16}$$

En otras palabras, las componentes de un vector covariante tranforman de la misma manera que los vectores de base, pero de manera opuesta a las componentes de un vector contravariante.[3]

También aquí podemos tomar esta propiedad como la definición práctica de un vector covariante: Cualquier objeto con N componentes y_i que bajo un cambio de base (4.4) transforma como (4.16) se le llama un *vector covariante*. Nótese que, debido al hecho de que anotamos el vector contravariante x^i con índice arriba y el vector covariante x_i con índice abajo, las reglas de transformación son (ligeramente) diferentes para cada caso.

De la regla de transformación de los vectores covariantes (4.16) podemos deducir las reglas de transformación de los vectores de la base dual. Con un argumento parecido (pero inverso) que en el caso de los vectores contravariantes vemos que la relación entre los vectores de la base dual viene dada por

$$\langle e'^i| = M^i{}_j\langle e^j|. \tag{4.17}$$

Una consecuencia directa de las reglas de transformación de la base y la base dual es que el producto escalar (4.12) entre vectores de base se conserva,

$$\langle e'^i|e'_j\rangle \;=\; M^i{}_k(M^{-1})^l{}_j\langle e^k|e_l\rangle \;=\; M^i{}_k(M^{-1})^l{}_j\delta^k{}_l \;=\; M^i{}_k(M^{-1})^k{}_j \;=\; \delta^i{}_j. \tag{4.18}$$

Y gracias a la linealidad del producto escalar, también se conserva el producto escalar entre dos vectores $\langle y|$ y $|x\rangle$, independientemente de la base en que se calcula. Efectivamente en componentes vemos que

$$\langle y|x\rangle \;=\; y'_i x'^i \;=\; (M^{-1})^k{}_i\, y_k\, M^i{}_l\, x^l \;=\; \delta^k{}_l\, y_k x^l \;=\; y_k x^k. \tag{4.19}$$

Esto es lo que uno esperaría, puesto que el producto escalar es un número real, cuyo valor no depende de la elección de base.

4.3. La métrica y las transformaciones ortogonales

En la sección anterior hemos introducido $\mathbb{R}^N$ como un espacio vectorial, con los vectores contravariantes como elementos y hemos definido el espacio dual con los vectores covariantes. Hemos visto cómo transforma cada uno bajo cambios de base y definido un producto escalar entre vectores covariantes y contravariantes que es independiente de la elección de base.

Sin embargo, esta estructura todavía es bastante pobre. No podemos dar una interpretación (física) al producto escalar, aparte de una aplicación lineal resultando en un número real. Tampoco podemos calcular la norma de un vector, o el ángulo entre dos vectores en $\mathbb{R}^N$. Para saber por ejemplo si una base $\{|e_i\rangle\}$ es ortogonal y normalizada, tenemos que calcular el producto escalar entre dos vectores de la base $|e_i\rangle$ y $|e_j\rangle$, mientras el producto escalar (4.12) y (4.14) sólo está definido entre un vector covariante y uno contravariante.

[3] De hecho, los nombre covariante y contravariante vienen de allí. En 1853, en su estudio de invariantes algebráicas, el matematico inglés James Joseph Sylvester (1814 - 1897) denominó las aplicaciones lineales y los vectores de $\mathbb{R}^N$ respectivamente vectores covariantes y contravariantes, por su propiedad de transformar junto con o en dirección opuesta a la base.

Lo que necesitamos claramente es una estructura matemática nueva que nos permita relacionar un vector contravariante $|x\rangle$ unívocamente con su correspondiente vector covariante $\langle x|$. Esto es lo que va hacer la métrica y gracias a ella podremos definir normas, ángulos y distancias en $\mathbb{R}^N$.

La *métrica* g_{ij} se puede ver como una operación que, actuando sobre un vector covariante, da un vector contravariante. En los vectores de base actúa como

$$|e_i\rangle = g_{ij}\langle e^j|. \tag{4.20}$$

Inversamente la *métrica inversa* g^{ij} transforma un vector covariante en uno contravariante

$$\langle e^i| = g^{ij}|e_j\rangle. \tag{4.21}$$

Nótese que la unicidad de la descomposición de vectores en una base implica que g_{ij} y g^{ij} están relacionados a través de la relación $g_{ij}g^{jk} = \delta^k{}_i$ (ejerc.). De allí que g^{ij} se llame la métrica inversa.

En general podemos relacionar las componentes del vector contravariante $|x\rangle$ con las componentes del vector covariante $\langle x|$ a través de la métrica de la siguiente manera: en la base $\{|e_i\rangle\}$, la descomposición del vector $|x\rangle$ es

$$x^i|e_i\rangle \;=\; x^i g_{ij}\langle e^j|, \tag{4.22}$$

donde hemos utilizado la relación (4.21) entre las dos bases. Ahora, también podemos pensar en el lado derecho de (4.22) como la descomposición de $\langle x|$ en la base $\{\langle e^i|\}$, lo que implica que las componentes x_i de $\langle x|$ se pueden escribir en función de las x^i como

$$x_i = g_{ij}x^j. \tag{4.23}$$

Con un argumento similar podemos también invertir esta relación, escribiendo las componentes x_i en función de las x^i (ejerc.)

$$x^i = g^{ij}x_j. \tag{4.24}$$

Por lo tanto vemos que la métrica y la inversa "*suben y bajan índices*", convirtiendo vectores covariantes en contravariantes y vice versa. Una vez introducida la métrica, se puede por lo tanto pensar en los vectores covariantes como un "truco matemático" para poder definir bien el producto escalar entre dos vectores contravariantes, puesto que en el fondo (4.23) y (4.24) están diciendo que un vector contravariante y su correspondiente vector covariante tienen la misma información física.

Para que las relaciones (4.23)-(4.24) sean consistentes con las reglas de transformación (4.5) y (4.16) de los vectores covariantes y contravariantes (o, equivalentemente, para que (4.21) y (4.20) sean consistentes con las reglas de transformación de los vectores de base), la métrica y su inversa tienen que transformar bajo un cambio de coordenadas M como

$$g'_{ij} = (M^{-1})^k{}_i(M^{-1})^l{}_j g_{kl}, \qquad\qquad g'^{ij} = M^i{}_k M^j{}_l\, g^{kl}, \tag{4.25}$$

donde g'_{ij} y g'^{ij} son la métrica y su inversa en las bases $\{|e'_i\rangle\}$ y $\{\langle e'^i|\}$. En general un cambio de coordenadas M puede hacer que la forma explícita de g'_{ij} sea muy diferente que la forma de g_{ij}. Sin embargo no hay que olvidar que las dos expresiones representan la misma métrica, la de $\mathbb{R}^N$, sólo en diferentes coordenadas. En la sección 4.4 veremos un ejemplo concreto.

La introducción de la métrica permite ampliar la definición del producto escalar (4.14), tal que también incluye el producto escalar entre dos vectores contravariantes. Utilizando la definición (4.21), tenemos que

$$|x\rangle\cdot|y\rangle \;=\; x^i y^k g_{ij}\langle e^j|e_k\rangle \;=\; g_{ij}x^i y^k\delta^j_k \;=\; g_{ij}x^i y^j \;= x_i y^j. \tag{4.26}$$

Con este producto escalar podemos entonces definir la norma $\||x\rangle\|$ de un vector $|x\rangle$ como

$$\||x\rangle\|^2 \;\equiv\; \langle x|x\rangle = \; x_i x^i \;=\; g_{ij}x^i x^j, \tag{4.27}$$

Finalmente el ángulo θ entre dos vectores $|x\rangle$ y $|y\rangle$ se define como

$$\cos\theta = \frac{\langle x|y\rangle}{\||x\rangle\| \, \||y\rangle\|}, \tag{4.28}$$

de modo que dos vectores son ortogonales cuando el producto escalar es cero. Dejamos como ejercicio demostrar que las reglas de transformación son tales que (4.26), (4.27) y (4.28) tienen el mismo valor en todos las bases.

Con estas definiciones podemos construir una base ortonormal en $\mathbb{R}^N$. Decimos que la base $\{|e_i\rangle\}$ es ortogonal y normalizada si

$$|e_i\rangle \cdot |e_j\rangle = \delta_{ij}, \tag{4.29}$$

donde δ_{ij} es la métrica de $\mathbb{R}^N$ en coordenadas cartesianas, a veces llamada por simplicidad (aunque no del todo correcto) la *métrica euclídea*,[4]

$$\delta_{ij} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}. \tag{4.30}$$

La definición (4.29) es consistente con (4.21), puesto que

$$|e_i\rangle \cdot |e_j\rangle \;=\; \delta_{ik}\langle e^k|e_j\rangle \;=\; \delta_{ik}\,\delta^k{}_j \;=\; \delta_{ij}. \tag{4.31}$$

La gran ventaja de utilizar una base ortonormal en $\mathbb{R}^N$ es que la forma (4.30) de la métrica es extremadamente simple. También está claro que la expresión explícita de las componentes de un vector covariante son idénticas a las componentes del vector contravariante correspondiente,

$$x^i = \delta^{ij}x_j, \qquad\qquad x_i = \delta_{ij}x^j, \tag{4.32}$$

una cosa que en una base general (y por lo tanto con una métrica general g_{ij}) no necesariamente es el caso, como demuestra (4.23)-(4.24). De igual manera, en una base ortonormal, las expresiones para el producto escalar y la norma toman una forma más sencilla:

$$\begin{aligned} \langle x|y\rangle \;&=\; x_iy^i \;=\; \delta_{ij}x^iy^j, \\ \||x\rangle\|^2 \;&=\; x_ix^i \;=\; \delta_{ij}x^ix^j, \end{aligned} \tag{4.33}$$

Valoraremos tanto la forma sencilla de la métrica en coordenadas cartesianas, que nos limitaremos para el resto de capítulo a trabajar exclusivamente con bases ortonormales. Sin embargo es útil enfatizar que todas las propiedades demostradas en en resto del capítulo siguen siendo verdad en bases arbitrarias, cambiando δ_{ij} por g_{ij}.

La norma (4.27) nos permite introducir el concepto de distancia en $\mathbb{R}^N$, ya que la distancia entre dos puntos x y y es la norma del vector $|x-y\rangle = |x\rangle - |y\rangle$. Está claro entonces que la expresión (4.33) para la norma no es más que la generalización a N dimensiones del Teorema de Pythagoras. En general la métrica resume las propiedades geométricas del espacio. Concretamente en este caso, (4.30) nos está diciendo que la geometría de $\mathbb{R}^N$ es la geometría plana de Euclides. Por futuro convenio introducimos el *elemento de línea* ds, la distancia entre un punto $\vec{x}$ y un punto cercano $\vec{x} + d\vec{x}$. Para el caso de $\mathbb{R}^N$ tenemos en coordenadas cartesianas que

$$ds^2 \;=\; \delta_{ij}dx^idx^j \;=\; (dx^1)^2 + (dx^2)^2 + ... + (dx^N)^2. \tag{4.34}$$

[4]Ojo, no se puede confundir la métrica euclídea δ_{ij} con la matriz identidad $\delta^i{}_j$. La métrica es un objeto (un tensor) que describe las propiedades geométricas del espacio, mientras la matriz identidad es la transformación trivial, la unidad. La diferencia se hará más clara en la siguiente sección.

En general un cambio de coordenadas M no transforma una base ortonormal en otra ortonormal, de modo que en general la métrica no preserva la forma (4.30) bajo un cambio de base M. Sin embargo sí existe un clase de transformaciones que tienen la propiedad que dejan la forma de la métrica invariante. Consideremos la clase de matrices M que tienen la propiedad que

$$(M^{-1})^l{}_j \, \delta_{kl} \;=\; M^l{}_k \, \delta_{lj}. \tag{4.35}$$

Si $\{|e'_i\rangle\}$ es una base ortonormal (es decir, la métrica toma la forma δ_{ij}), entonces, utilizando (4.25) es fácil ver que la forma de la métrica en la nueva base es

$$g'_{ij} \;=\; (M^{-1})^k{}_i \, M^l{}_k \, \delta_{lj} \;=\; \delta_{ij}. \tag{4.36}$$

En otras palabras, en la nueva base la métrica también es de la forma (4.30), puesto que las transformaciones consideradas dejan su forma invariante:

$$\delta_{ij} \;=\; (M^{-1})^k{}_i (M^{-1})^l{}_j \delta_{kl}. \tag{4.37}$$

Las transformaciones que satisfacen la condición (4.35) se llaman *transformaciones ortogonales*. Tienen la buena propiedad de que transforman una base ortonormal en otra base ortonormal (esta es la razón por que deja la métrica invariante).

Hay algunas propiedades más de las matrices ortogonales que merecen la pena mencionar. Si multiplicamos ambos lados de la condición (4.35) condición de ortogonalidad con δ^{ik}, tenemos que

$$(M^{-1})^i{}_j = \delta^{ik} M^l{}_k \delta_{lj}. \tag{4.38}$$

La combinación de métricas contraída de esta manera con $M^i{}_j$ se conoce como la *traspuesta* M^T de la matriz M,

$$(M^T)^i{}_j = \delta^{ik} M^l{}_k \delta_{lj}. \tag{4.39}$$

En otras palabras, la misma definición de las transformaciones ortogonales nos está diciendo que la inversa de la matriz ortogonal M es su traspuesta (en forma matricial $M^T = M^{-1}$). Además las matrices ortogonales M son reales y tienen determinante ± 1. Esta última propiedad se saca directamente de (4.38), ya que si $M^T = M^{-1}$, tomando el determinante vemos que $|M| = |M|^{-1}$, o en otras palabras, $|M| = \pm 1$.

Desde en punto de vista geométrico, las transformaciones ortogonales son las rotaciones (si tienen determinante 1) y las reflexiones (si tienen determinante -1). Las transformaciones ortogonales en N dimensiones forman el grupo $O(N)$, que contiene el grupo de las rotaciones $SO(N)$ como subgrupo.

Para ser consistentes con nuestro propósito de trabajar sólo con bases ortonormales, está claro que para el resto de este capítulo nos limitaremos a cambios de base que son transformaciones ortogonales. Sobra decir, sin embargo, que todas las propiedades mencionadas a partir de aquí siguen siendo válidas para transformaciones lineales arbitrarias.

4.4. Ejemplo concreto

A este punto conviene ilustrar algunas de las propiedades estudiadas con unos ejemplos concretos. Consideramos el cambio de coordenadas en $\mathbb{R}^2$, parametrizado por la matriz

$$(M^{-1})^i{}_j = \begin{pmatrix} \frac{a}{\sqrt{2}} & -\frac{a}{\sqrt{2}} \\ \frac{b}{\sqrt{2}} & \frac{b}{\sqrt{2}} \end{pmatrix} \qquad \Longleftrightarrow \qquad M^i{}_j = \begin{pmatrix} \frac{1}{\sqrt{2}a} & \frac{1}{\sqrt{2}b} \\ -\frac{1}{\sqrt{2}a} & \frac{1}{\sqrt{2}b} \end{pmatrix} \tag{4.40}$$

con a y b números reales arbitrarios. Tenemos que $\det(M^{-1}) = ab$, de modo que en general M^{-1} no es una transformación ortogonal. Sin embargo, para $a = b = 1$, M^{-1} se reduce a una rotación

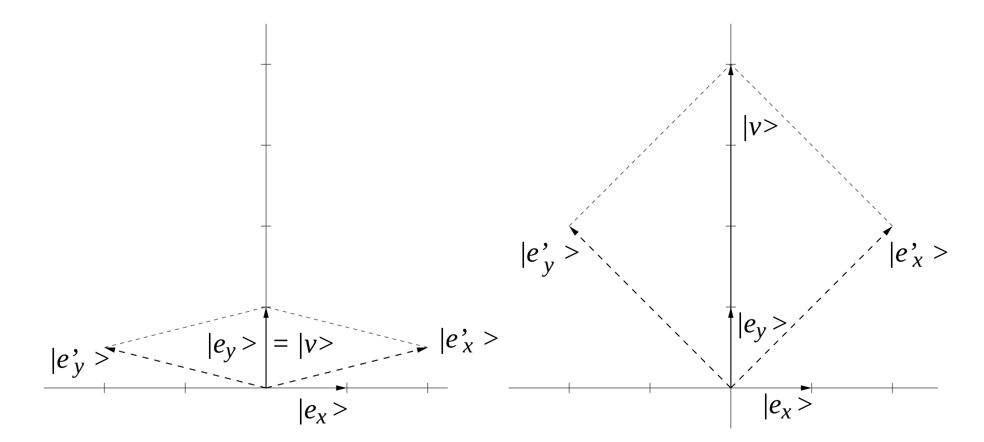


Figura 4.2: *El cambio de base (4.40) con $a = 2\sqrt{2}$ y $b = \sqrt{2}/2$ (izquierda) y $a = b = 2\sqrt{2}$ (derecha). En el segundo caso, una base ortonormal transforma en otra base ortogonal, pero no normalizada.*

bajo un ángulo $\frac{\pi}{4}$ en sentido contrarreloj. Por otro lado, si $a = b \neq 1$, M^{-1} transforma una base ortonormal en una base ortogonal, pero no normalizada, puesto que aparte de una rotación con ángulo $\frac{\pi}{4}$, escalea los vectores de base con un factor a^2.

Sea ahora $\{|e_i\rangle\}$ una base ortonormal, es decir

$$|e_x\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad |e_y\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{4.41}$$

Entonces la nueva base $\{|e_i'\rangle\}$ viene dada en función de los vectores de la base vieja como (4.4)

$$|e_x'\rangle = \frac{a}{\sqrt{2}}|e_x\rangle + \frac{b}{\sqrt{2}}|e_y\rangle, \qquad |e_y'\rangle = -\frac{a}{\sqrt{2}}|e_x\rangle + \frac{b}{\sqrt{2}}|e_y\rangle, \tag{4.42}$$

y vice versa

$$|e_x\rangle = \frac{1}{\sqrt{2}a}|e_x'\rangle - \frac{1}{\sqrt{2}a}|e_y'\rangle, \qquad |e_y\rangle = \frac{1}{\sqrt{2}b}|e_x'\rangle + \frac{1}{\sqrt{2}b}|e_y'\rangle. \tag{4.43}$$

Un vector $|v\rangle$, que en la base $\{|e_i\rangle\}$ tiene componentes $|v\rangle = \sqrt{2}b|e_y\rangle$ tiene en la nueva base la forma $|v\rangle = |e_x'\rangle + |e_y'\rangle$, puesto que por (4.5)

$$\begin{pmatrix} \frac{1}{\sqrt{2}a} & \frac{1}{\sqrt{2}b} \\ -\frac{1}{\sqrt{2}a} & \frac{1}{\sqrt{2}b} \end{pmatrix} \begin{pmatrix} 0 \\ \sqrt{2}b \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}. \tag{4.44}$$

Una aplicación lineal $\langle w| = \langle e^x| + 2\langle e^y|$ tendrá en la base $\{\langle e_i'|\}$ componentes

$$\langle w| \;=\; \frac{a+2b}{\sqrt{2}}\langle e'^x| \;+\; \frac{-a+2b}{\sqrt{2}}\langle e'^y|, \tag{4.45}$$

o bien por (4.16), o bien por la la multiplicación de matrices

$$\begin{pmatrix} 1 & 2 \end{pmatrix} \begin{pmatrix} \frac{a}{\sqrt{2}} & -\frac{a}{\sqrt{2}} \\ \frac{1}{\sqrt{2}b} & \frac{b}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} \frac{a+2b}{\sqrt{2}} & \frac{-a+2b}{\sqrt{2}} \end{pmatrix}. \tag{4.46}$$

El producto escalar entre los vectores $\langle w|$ y $|v\rangle$ tiene el mismo valor en las dos bases $\{|e_i\rangle\}$ y $\{|e_i'\rangle\}$, ya que

$$\begin{aligned} \langle w|v\rangle &= w_i v^i \;=\; 0 + 2\sqrt{2}b \;=\; 2\sqrt{2}b \\ &= w_i' v'^i \;=\; \frac{a+2b}{\sqrt{2}} + \frac{-a+2b}{\sqrt{2}} \;=\; 2\sqrt{2}b. \end{aligned} \tag{4.47}$$

Dado que $\{|e_i\rangle\}$ es una base ortonormal en $\mathbb{R}^2$, la métrica en esta base es la euclídea δ_{ij}. Sin embargo la forma de la métrica en la base $\{|e'_i\rangle\}$ se calcula con (4.25), siendo de la forma

$$g_{ij} = \begin{pmatrix} \frac{a^2+b^2}{2} & \frac{b^2-a^2}{2} \\ \frac{b^2-a^2}{2} & \frac{a^2+b^2}{2} \end{pmatrix}. \qquad (4.48)$$

Las componentes del vector covariante $\langle v|$, correspondiente a $|v\rangle$, en las bases $\{\langle e^i|\}$ y $\{\langle e'^i|\}$ vienen dadas respectivamente por (ejerc.)

$$v_x = 0, \qquad v_y = \sqrt{2}b, \qquad v'_x = b^2, \qquad v'_y = b^2, \qquad (4.49)$$

de modo que la norma $\||v\rangle\|$ en ambas bases viene dado por

$$\begin{aligned} \||v\rangle\|^2 &= v_i v^i \;=\; 0 + (\sqrt{2}b)^2 \;=\; 2b^2 \\ &= v'_i v'^i \;=\; b^2 + b^2 \;=\; 2b^2. \end{aligned} \qquad (4.50)$$

Por lo tanto, la norma de vectores también es independiente de la base en que se calcula, como es de esperar, por lo menos siempre y cuando se use la misma medida en ambas bases.

Miremos esto último en un poco más de detalle: el resultado calculado arriba implica que los observadores $\mathcal{O}$ y $\mathcal{O}'$, que usan los sistemas de referencia $\{|e_i\rangle\}$ y $\{|e'_i\rangle\}$ respectivamente, estarán de acuerdo sobre la longitud de vectores, si los dos están de acuerdo en usar como unidad de longitud la norma $\||e_x\rangle\|$ del vector de base de $|e_x\rangle$. En este caso, las reglas de transformación de vectores co- y contravariantes, de la métrica y del producto escalar les dicen cómo relacionar los datos de ambos observadores.

Sin embargo, en la práctica $\||e_x\rangle\|$ no es la unidad de longitud más natural para el observador $\mathcal{O}'$. Esto está claro en el caso donde $a = b \neq 1$ (véase Figura 4.2), donde la base $\{|e'_i\rangle\}$ también es ortogonal. En este caso, $\mathcal{O}'$ cogerá de manera natural $\||e'_x\rangle\|$ como unidad de longitud y usará la métrica euclídea δ_{ij} como métrica en su sistema de referencia y encontrará que según él las distancias son un factor a^2 $(= b^2)$ más cortas de lo que afirma $\mathcal{O}$. Este efecto ocurre en general, también para bases no-ortogonales y/o bases con vectores con distintas normas (según $\mathcal{O}'$). Concretamente $\mathcal{O}'$ afirmaría que el vector $|v\rangle$ tiene norma $\||v\rangle\| = \sqrt{2}$ en su sistema de referencia, tanto en el caso $a = b$ como $a \neq b$.

En general los observadores $\mathcal{O}$ y $\mathcal{O}'$ sólo estarán de acuerdo sobre las distancias en ambos sistemas de referencias si los dos utilizan bases cartesianas que estén normalizadas de la misma manera (es decir, si usan la misma unidad de longitud). En otras palabras, sólo estarán de acuerdo si el cambio de coordenadas M^{-1} transforma la base ortonormal $\{|e_i\rangle\}$ en otra base ortonormal $\{|e'_i\rangle\}$, y por lo tanto deja la métrica euclídea invariante. Como hemos visto en la sección anterior, son precisamente las transformaciones ortogonales (4.35) las que tienen esta propiedad. Aunque en general todas las transformaciones preservan el producto escalar y la norma en el sentido de (4.47) y (4.50), se suele decir que sólo las transformaciones ortogonales preservan las distancias.

4.5. Álgebra de tensores

En la sección anterior hemos definido los vectores contravariantes x^i y covariantes x_i, que bajo cambios de coordenadas transforman como

$$x'^i = M^i{}_j x^j, \qquad x'_i = (M^{-1})^j{}_i x_j, \qquad (4.51)$$

y que están relacionados a través de la métrica δ_{ij} y la métrica inversa δ^{ij} como

$$x_i = \delta_{ij} x^j, \qquad x^i = \delta^{ij} x_j. \qquad (4.52)$$

Con estas definiciones (4.51), se puede construir objetos más generales: los escalares y los tensores.

Un *escalar* ϕ es un objeto que tiene el mismo valor en todos los sistemas de referencia

$$\phi' = \phi. \tag{4.53}$$

En otras palabras, un escalar es un invariante bajo cambios de coordenadas. Un ejemplo de un escalar es el producto escalar de dos vectores, como hemos visto previamente. Del mismo modo, una función escalar (o un campo escalar) es una función $\phi(x^i)$ que asigna a cada punto $\vec{x}$ un valor numérico $\phi(x^i)$. Dado que es una función escalar, ϕ es invariante bajo cambios de coordenadas:

$$\phi'(x'^i) = \phi(x^i). \tag{4.54}$$

En general ϕ será una función de x^i diferente que ϕ' de x'^i, pero los dos asignarán el mismo valor numérico al mismo punto del espacio.

Considérese ahora el objeto T^{ij} construido de los vectores contravariantes a^i y b^j a través del *producto exterior* o el *producto tensorial*: $T^{ij} = a^i b^j$. Matemáticamente hablando es un objeto que vive en el espacio $\mathbb{R}^N \times \mathbb{R}^N$, pero lo que realmente nos interesa es cómo transforma bajo un cambio de base. De la definición de los vectores, está claro que T^{ij} transforma como

$$T'^{ij} \;=\; a'^i b'^j \;=\; M^i{}_k a^k M^j{}_l b^l \;=\; M^i{}_k M^j{}_l T^{kl}. \tag{4.55}$$

Definimos un *tensor contravariante de rango 2* como cualquier objeto de N^2 componentes que transforma como (4.55). Obsérvese que un producto tensorial de dos vectores contravariantes, como T^{ij}, siempre es un tensor contravariante, pero no todos los tensores contravariantes de rango 2 se pueden escribir como el producto tensorial de dos vectores contravariantes.

Del mismo modo podemos definir un *tensor covariante de rango 2* como un objeto T_{ij} de ${}^*\mathbb{R}^N \times {}^*\mathbb{R}^N$, que transforma como

$$T'_{ij} \;=\; (M^{-1})^k{}_i (M^{-1})^l{}_j T_{kl}, \tag{4.56}$$

y en general un *tensor mixto de rango m contravariante y rango n covariante* (o simplemente *un tensor de rango (m,n)*) un objeto de $\mathbb{R}^N \times \ldots \times \mathbb{R}^N \times {}^*\mathbb{R}^N \times \ldots \times {}^*\mathbb{R}^N$ que transforma como como

$$T'^{i_1\ldots i_m}{}_{j_1\ldots j_n} \;=\; M^{i_1}{}_{k_1}\ldots M^{i_m}{}_{k_m} (M^{-1})^{l_1}{}_{j_1}\ldots (M^{-1})^{l_n}{}_{j_n} T^{k_1\ldots k_m}{}_{l_i\ldots l_n}. \tag{4.57}$$

En otras palabras, cada índice co(ntra)variante transforma como si fuera el índice de un vector co(ntra)variante.

Comparando esta definición con las reglas de transformación (4.25) de la métrica, vemos que la métrica g_{ij} es en realidad un tensor covariante de rango 2. De la misma manera se puede pensar en un vector co(ntra)variante como un tensor de rango 1 y un escalar como un tensor de rango 0.

Por último, es útil introducir los siguientes tensores importantes: la *delta de Kronecker*, definida en (4.13) y el *tensor de Levi-Civita* $\varepsilon^{i_1\cdots i_N}$, definida como

$$\varepsilon^{i_1\ldots i_N} \;=\; \begin{cases} 1 & \text{cuando } (i_1 \ldots i_N) \text{ es una permutación par de } (12\ldots N), \\ -1 & \text{cuando } (i_1 \ldots i_N) \text{ es una permutación impar de } (12\ldots N), \\ 0 & \text{en todos los demás casos.} \end{cases} \tag{4.58}$$

Por construcción la delta de Kronecker es un tensor simétrico. La delta es en realidad la variante $(1,1)$ de la métrica de $\mathbb{R}^N$, es decir $\delta^i{}_j$ es lo mismo que δ_{ij} con un índice subido:

$$\delta^i{}_j = \delta^{ik}\delta_{kj}. \tag{4.59}$$

La delta de Kronecker es un *tensor fundamental*, lo que quiere decir que mantiene la misma forma en cualquier sistema de referencia:

$$\delta'^i{}_j = M^i{}_k (M^{-1})^l{}_j \, \delta^k{}_l \;=\; \delta^i{}_j. \tag{4.60}$$

Insistimos otra vez que no se debería confundir la delta de Kronecker con la matriz identidad (4.7): la matriz identidad es una transformación (trivial), mientras la delta es un tensor que transforma bajo transformaciones en $\mathbb{R}^N$.

Por construcción, el tensor de Levi-Civita es completamente antisimétrico en todos sus índices. Eso hace que estrictamente hablando no sea un tensor, puesto que en sus reglas de transformación aparecen el signo y el valor absoluto del determinante $(\det M^{-1})$ del cambio de coordenadas $(M^{-1})^i{}_j$:

$$\varepsilon'^{i_1...i_N} = \text{sgn}(\det M^{-1})\,|\det M^{-1}|\, M^{i_1}{}_{j_1}...M^{i_N}{}_{j_N}\varepsilon^{j_1...j_N}. \tag{4.61}$$

Objetos que transforman con el signo del determinante de la transformación se llaman *pseudo-tensores* y objetos que transforman con el valor absoluto del determinante de la transformación (a la potencia p) se llaman *densidades tensoriales (de peso p)*. Estrictamente hablando el tensor de Levi-Civita es por lo tanto una pseudo-tensor densidad de peso 1.[5]

Como la delta Kronecker, también el tensor de Levi-Civita es un tensor fundamental. De la regla de transformación (4.61) deducimos que

$$\begin{aligned} \varepsilon'^{i_1...i_N} &= \text{sgn}(\det M^{-1})\,|\det M^{-1}|\, M^{i_1}{}_{j_1}...M^{i_N}{}_{j_N}\varepsilon^{j_1...j_N} \\ &= \det M^{-1}\ \det M\ \varepsilon^{i_1...i_N} \ = \ \varepsilon^{i_1...i_N}, \end{aligned} \tag{4.62}$$

donde en la segunda igualdad hemos utilizado la identidad

$$\det M\ \varepsilon^{i_1...i_N} = M^{i_1}{}_{j_1}...M^{i_N}{}_{j_N}\varepsilon^{j_1...j_N}, \tag{4.63}$$

que se demuestra fácilmente de la propia definición del determinante

$$\det M = \tfrac{1}{N!}\, M^{i_1}{}_{j_1}...M^{i_N}{}_{j_N}\ \varepsilon_{i_1...i_N}\ \varepsilon^{j_1...j_N}. \tag{4.64}$$

4.6. Operaciones con tensores

Una primera propiedad de los tensores es una generalización directa de una propiedad conocida de vectores (tensores): de las reglas de transformación (4.51) ((4.57)) está claro que si una componente de un vector (tensor) es cero en una base, en general no lo será en otra base. Sin embargo, también está claro que si todas las componentes de un vector (tensor) son cero en una base, lo serán en todas las bases.

De la regla de transformación (4.57) también está claro que si $A^{i\cdots}{}_{j\ldots}$ y $B^{i\cdots}{}_{j\ldots}$ son dos tensores de rango (m, n), la combinación lineal

$$C^{i_1...i_m}{}_{j_1...j_n} = \alpha A^{i_1...i_m}{}_{j_1...j_n} + \beta B^{i_1...i_m}{}_{j_1...j_n} \tag{4.65}$$

también es un tensor de rango (m, n).

Si un tensor tiene la propiedad que $T^{ij\cdots}{}_{\ldots} = T^{ji\cdots}{}_{\ldots}$, decimos que T es un *tensor simétrico* en i y j. Si tiene la propiedad que $T^{ij\cdots}{}_{\ldots} = -T^{ji\cdots}{}_{\ldots}$, decimos que es un *tensor anti-simétrico* en i y j. Utilizando (4.65) se deriva directamente que si un tensor es (anti-)simétrico en una base, lo es en todas.

De la forma en que hemos introducido los tensores también está claro que el producto tensorial de un tensor de (m, n) con uno de rango (p, q)

$$C^{i_1...i_m}{}_{j_1...j_n}{}^{k_1...k_p}{}_{l_1...l_q} = A^{i_1...i_m}{}_{j_1...j_n} B^{k_1...k_p}{}_{l_1...l_q} \tag{4.66}$$

[5] El estudiante atento habrá observado que al ser $(M^{-1})^i{}_j$ una transformación ortogonal, tenemos que $|\det M^{-1}| = 1$. Sin embargo preferimos poner el factor $|\det M^{-1}|$ explícitamente por razones pedagógicas, para demostrar que el tensor de Levi-Civita es un pseudo-tensor bajo las transformaciones ortogonales, pero una densidad pseudo-tensorial bajo transformaciones más generales.

es un tensor de rango $(m+p, n+q)$.

Se puede cambiar el rango de los tensores utilizando la métrica para "bajar índices". Siendo $A^{i_1\ldots i_m}{}_{j_1\ldots j_n}$ un tensor de rango (m, n), entonces

$$C^{i_1\ldots i_{m-1}}{}_{kj_1\ldots j_n} = \delta_{kl} A^{i_1\ldots i_{m-1}l}{}_{j_1\ldots j_n} \tag{4.67}$$

es un tensor de rango $(m-1, n+1)$. De la misma manera se puede "subir índices" con la métrica inversa δ^{ij} y convertir un tensor de rango (m, n) en uno de rango $(m+1, n-1)$. Un ejemplo de esto ya lo hemos visto en (4.52), al cambiar un vector covariante por uno contravariante.

Se puede bajar el rango de un tensor contrayendo un índice covariante con uno contravariante. Si $A^{i_1\ldots i_m}{}_{j_1\ldots j_n}$ es un tensor de rango (m, n), entonces

$$C^{i_1\ldots i_{m-1}}{}_{j_1\ldots j_{n-1}} = A^{i_1\ldots i_{m-1}k}{}_{j_1\ldots j_{n-1}k} \tag{4.68}$$

es un tensor de rango $(m-1, n-1)$. Lo mismo se puede hacer contrayendo dos índices del mismo tipo con la métrica o la inversa. Por ejemplo

$$D^{i_1\ldots i_{m-2}}{}_{j_1\ldots j_n} = \delta_{kl} A^{i_1\ldots i_{m-2}kl}{}_{j_1\ldots j_n} \tag{4.69}$$

es un tensor de rango $(m-2, n)$. Visto así, el producto escalar es un tensor de rango 0 (un escalar) que surge de la contracción del producto tensorial de dos vectores $x_i y^j$. Otro ejemplo es que la traza $T^i{}_i = \delta_{ij} T^{ij} = \delta^{ij} T_{ij}$ de un tensor de rango 2 es un escalar. Esto explica el conocido resultado de que la traza tiene el mismo valor en todas las bases.

El operador ∂, la derivada con respecto a las coordenadas, se comporta como un vector. Para ver esto, es mejor mirar como actúa sobre un campo escalar. Consideramos la derivada parcial $\partial V/\partial x^i$. Por la regla de la cadena tenemos que

$$\frac{\partial V}{\partial x^i} = \frac{\partial V}{\partial x'^k}\frac{\partial x'^k}{\partial x^i} = \frac{\partial V}{\partial x'^k} M^k{}_i, \tag{4.70}$$

donde en la última igualdad hemos utilizado la relación (4.5) entre las coordenadas x^i y x'^i. En otras palabras la derivada $\partial_i V \equiv \partial V/\partial x^i$ tiene la misma regla de transformación que un vector covariante:

$$\partial'_i V = (M^{-1})^j{}_i \partial_j V. \tag{4.71}$$

Del mismo modo se puede demostrar que la derivada $\partial^i V \equiv \partial V/\partial x_i$ transforma como un vector contravariante. En notación vectorial muchas veces se utiliza el operador *nabla* $\vec{\nabla} = (\partial^i.)\vec{e}_i$ y se identifica con el operador matemático de la derivada direccional.

Tomando en cuenta el hecho que ∂_i y ∂^i se comportan como un vector covariante y contravariante respectivamente, se puede utilizar estos operadores para actuar sobre campos tensoriales. Dado que la derivada de un campo escalar es un vector, en general la derivada ∂^i de un campo tensorial de rango (m, n) es un tensor de rango $(m+1, n)$ y la derivada ∂_i de un campo tensorial de rango (m, n) es un tensor de rango $(m, n+1)$.

De este modo se puede interpretar los conocidos operadores diferenciales en $\mathbb{R}^3$ (la mayoría fácilmente generalizables a N dimensiones) en términos de (contracciones de) vectores y tensores: el *gradiente* $\vec{\nabla}\phi = \partial^i\phi\,\vec{e}_i$ de un campo escalar ϕ es un vector, al ser la derivada de un escalar. La *divergencia* $\vec{\nabla}\cdot\vec{A} = \partial_i A^i$ de un vector contravariante A^i es un escalar, al ser el producto escalar entre los vectores ∂_i y A^i. El *laplaciano* $\Delta\phi = \partial_i\partial^i\phi$ de un campo escalar ϕ es un escalar por ser el producto escalar entre los vectores ∂_i y $\partial^i\phi$. Finalmente, el *rotacional* $\vec{\nabla}\times\vec{A} = \varepsilon^{ijk}\partial_j A_k$ de un vector covariante es un vector contravariante, por ser la contracción del tensor de Levi-Civita tridimensional ε^{ijk} con el tensor de rango (0, 2) $\partial_i A_j$. Esta operación corresponde al producto vectorial entre los vectores ∂_i y A^i y solamente está definido en tres dimensiones, porque sólo en tres dimensiones el tensor de Levi-Civita ε^{ijk} puede convertir un tensor antisimétrico de rango 2 en un vector.

La gran ventaja del uso de tensores es que si una relación tensorial del tipo

$$A^{i_1 \ldots i_m}{}_{j_1 \ldots j_n} \; = \; B^{i_1 \ldots i_m}{}_{j_1 \ldots j_n} \; + \; C^{i_1 \ldots i_m} D_{j_1 \ldots j_n} \tag{4.72}$$

es válida en una base, también lo será en cualquier otra base, puesto que los dos lados de la ecuación transforman bajo un cambio de base de la misma manera. Un ejemplo de esta propiedad hemos visto ya en el Capítulo 2, al demostrar la covariancia de la Segunda Ley de Newton bajo rotaciones. Por lo tanto, el hecho de que por el Principio de la Relatividad las leyes de la física no dependan del sistema de referencia utilizado, implica que estas leyes deberían escribirse en término de objetos que transforman bien bajo cambios de coordenadas, es decir en términos de escalares, vectores y tensores. En otras palabras, la estructura matemática restringe la forma de las leyes a ecuaciones vectoriales y tensoriales. Este principio se llama el *Principio de Covariancia* y veremos que forma la base del formalismo matemático de la teoría de la relatividad.

Capítulo 5

Relatividad especial en formulación covariante

A partir de ahora el espacio en sí mismo y el tiempo en sí mismo están destinados a diluirse en meras sombras y sólo un tipo de unión de los dos conservará una realidad independiente.
(H. Minkowski, 1908)

Hasta ahora apenas hemos hecho algo nuevo: en el Capítulo 3 hemos repasado la relatividad especial y en el Capítulo 4 un poco de álgebra elemental. En este capítulo combinaremos las técnicas de Capítulo 4 con la física de Capítulo 3 para llegar a una formulación de la relatividad especial que sea independiente de los observadores. Aunque estrictamente hablando tampoco en este capítulo haremos nada nuevo, la formulación introducida aquí nos será de extrema utilidad para el desarollo de la relatividad general.

A partir de este capítulo utilizaremos las llamadas *unidades naturales* , en las que pondremos $c = 1$. Una de las lecciones de este capítulo es que el tiempo se comporta como una coordenada más en el espaciotiempo y la mejor manera de ver esto es utilizar unidades naturales.[1]

5.1. El espacio de Minkowski (otra vez)

Discutiremos ahora la estructura del espacio y el tiempo y sus propiedades algebráicas y geométricas. La imagen del espacio y el tiempo que surge de la relatividad especial es una donde ambos están íntimamente conectados y forman una estructura llamada el *espaciotiempo*. La dirección temporal es distinta a las espaciales, pero la distinción es mucho menos clara que en la física newtoniana. No hay una sola dirección temporal absoluta, sino que cada observador ve su propia dirección temporal, que es distinta de la de los demás.

Consideramos el espacio de Minkowski, un espacio continuo 4-dimensional, donde cada punto corresponde con un suceso, es decir un evento que ocurre en una posición (x, y, z) en un momento t en cierto sistema de coordenadas. Cada suceso está por lo tanto caracterizado por un

[1] Puede ser al principio un poco confuso que cantidades físicas (newtonianamente) distintas ahora tengan la misma dimensión, como tiempo y distancia, o energía, masa y momento. Pero expresar una distancia como un intervalo de tiempo o vice versa no tan es raro, incluso en la vida cotidiana. Decir que la distancia Madrid-Granada son 4 horas y media en coche es un ejemplo familiar. Para recuperar las dimensiones familiares sólo hace falta introducir en los sitios adecuados los factores c adecuados.

vector cuadrimensional

$$x^\mu = \begin{pmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \end{pmatrix} = \begin{pmatrix} t \\ x \\ y \\ z \end{pmatrix}. \tag{5.1}$$

Hemos visto en el Capítulo 3 que la norma $\|x^\mu\|$ de los vectores se define en el espacio de Minkowski como

$$\|x^\mu\|^2 \;=\; (x^0)^2 - (x^1)^2 - (x^2)^2 - (x^3)^2. \tag{5.2}$$

Del Capítulo 4 sabemos que con esta definición de la norma, hemos introducido implícitamente una métrica. Efectivamente, se puede escribir (5.2) como

$$\|x^\mu\|^2 \;=\; x_\mu x^\mu \;=\; \eta_{\mu\nu} x^\mu x^\nu, \tag{5.3}$$

donde $\eta_{\mu\nu}$ se llama la *métrica de Minkowski* y es de la forma[2]

$$\eta_{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} = \mathrm{diag}\,(1, -1, -1, -1). \tag{5.4}$$

El elemento de línea, que resume las propiedades geométricas del espaciotiempo viene por lo tanto dado por

$$ds^2 = \eta_{\mu\nu} dx^\mu dx^\nu = (dx^0)^2 - (dx^1)^2 - (dx^2)^2 - (dx^3)^2. \tag{5.5}$$

La métrica y su inversa nos permiten subir y bajar los índices y convertir vectores covariantes en contravariantes y vice versa,

$$x_\mu = \eta_{\mu\nu} x^\nu, \qquad\qquad x^\mu = \eta^{\mu\nu} x_\nu \tag{5.6}$$

de modo que el producto escalar se puede escribir como

$$x_\mu y^\mu \;=\; \eta_{\mu\nu} x^\mu y^\nu \;=\; \eta^{\mu\nu} x_\mu y_\nu \;=\; x^\mu y_\mu, \tag{5.7}$$

o, explícitamente en componentes

$$x_\mu y^\mu \;=\; x_0 y^0 + x_1 y^1 + x_2 y^2 + x_3 y^3 \;=\; x^0 y^0 - x^1 y^1 - x^2 y^2 - x^3 y^3. \tag{5.8}$$

Como vemos, hay que tener un particular cuidado al subir y bajar índices con los signos que aparecen por la métrica (5.4). Esto es una gran diferencia entre la geometría euclídea y la del espacio de Minkowski: las componentes de un vector cambian al pasar el vector de covariante a contravariante o vice versa. En particular, debido a los signos en la métrica (5.4), las componentes espaciales adquieren un signo. Si un vector contravariante tiene componentes

$$a^\mu = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}, \tag{5.9}$$

el vector covariante correspondiente tiene componentes $a_\mu = (1, -2, -3, -4)$. Si la métrica $g_{\mu\nu}$ tiene una forma más general,[3] la relación entre las componentes de vectores co- y contravariantes será más complicada.

[2] A veces la métrica se presenta con los signos opuestos (es decir $\eta_{\mu\nu} = \mathrm{diag}\,(-1, 1, 1, 1)$). El signo de la métrica es puro convenio y, aunque algunas fórmulas pueden tener signos diferentes, la física es por supuesto independiente del convenio utilizado. Para un resumen de los convenios que utilizamos en este curso, referimos a Apéndice A.

[3] Aquí referimos tanto a la métrica del espacio de Minkowski en coordenadas no-cartesianas, como a métricas de espacios más complejos.

La dirección temporal destaca en la métrica (5.4) porque lleva el signo contrario de las direcciones espaciales. Esto tiene como consecuencia directa que la norma no sea definida positiva, es decir: existen vectores cuya norma al cuadrado es negativa. Se dice que la métrica (5.4) tiene *signatura (3,1)*. El espacio de Minkowski por lo tanto no es euclídeo, sino *lorentziano*. La diferencia con el espacio euclídeo es la presencia de la dirección temporal que hace posible una dinámica.

Los vectores se llaman *vectores temporales* si su norma al cuadrado es positiva, *vectores nulos* si su norma es cero y *vectores espaciales* si la norma al cuadrado es negativa. Entre los vectores temporales distinguimos los *vectores temporales dirigidos hacia el futuro* y los *vectores temporales dirigidos hacia el pasado*, dependiendo de si el producto escalar entre el vector temporal y el vector de base $\vec{e}_t$ es positivo o negativo. Una curva se llama *curva temporal* si en cada punto de la curva el vector tangente es un vector temporal. De igual modo se llama *curva espacial (nula)* si en cada punto el vector tangente es espacial (nulo). En general una curva arbitraria puede ser temporal, nula o espacial por trozos.

El conjunto de todas las rectas nulas que pasan por un punto x^μ se llama *cono de luz en* x^μ. La estructura del cono de luz y la distinción entre vectores y curvas temporales, nulos y espaciales está bien definida, en el sentido de que es igual para todos los observadores. Esto es porque el grupo de Lorentz, que relaciona diferentes observadores, preserva esta estructura del espacio de Minkowski, como veremos en la siguiente sección.

5.2. El grupo de Lorentz

El *grupo de Lorentz* $\mathcal{L}$ se define como el grupo de transformaciones lineales y homogéneas[4]

$$x'^\mu = \Lambda^\mu{}_\nu x^\nu \tag{5.10}$$

que dejan la métrica (5.4) invariante,

$$\eta_{\mu\nu} = \Lambda^\rho{}_\mu \Lambda^\lambda{}_\nu \eta_{\rho\lambda}. \tag{5.11}$$

Comparando esta expresión con la regla de transformación (4.56), vemos que $\eta_{\mu\nu}$ transforma como un tensor de rango $(0,2)$ bajo las transformaciones de Lorentz (como era de esperar). Además, la definición (5.11) tiene mucha similitud con la fórmula (4.37) y con la definición (4.35) de las transformaciones ortogonales en $\mathbb{R}^N$. Efectivamente, de (5.11) se deriva fácilmente que

$$(\Lambda^{-1})^\lambda{}_\rho \ = \ \eta_{\rho\mu}\, \Lambda^\mu{}_\nu\, \eta^{\nu\lambda}. \tag{5.12}$$

Comparando esta definición con (4.38) y (4.39), vemos que de cierto modo esta fórmula nos está diciendo que la traspuesta de una transformación de Lorentz es su inversa, una propiedad de las transformaciones ortogonales, sólo que aquí la traspuesta está definida con respecto a la métrica de Minkowski $\eta_{\mu\nu}$. Esto es general: la única diferencia entre las transformaciones en el espacio de Minkowski y las transformaciones ortogonales en el caso euclídeo es que aquí aparece $\eta_{\mu\nu}$ en lugar de la métrica euclídea δ_{ij}. Podemos interpretar por lo tanto las transformaciones de Lorentz (5.11) como el grupo de transformaciones ortogonales en el espacio de Minkowski y las fórmulas (5.11) y (5.12) como los análogos lorentzianos de (4.37) y (4.35) en $\mathbb{R}^N$.

Del Capítulo 4 podemos derivar matemáticamente muchas propiedades que ya habíamos visto en el Capítulo 3. En particular sabemos que un vector covariante transforma con la matriz inversa de un vector contravariante

$$x'_\mu = (\Lambda^{-1})^\nu{}_\mu x_\nu, \tag{5.13}$$

[4]Hasta ahora nos hemos referido a una transformación de Lorentz como la que relaciona x y t como en (3.19). A veces a (3.19) se le llama una *transformación de Lorentz especial* o por el término en inglés, *boost* (empujón), para distinguirla de los demás elementos del grupo de Lorentz. A partir de ahora llamaremos una transformación de Lorentz a cualquier elemento de $\mathcal{L}$.

de modo que el producto escalar (y por lo tanto la norma) de vectores en el espacio de Minkowski se conserva. Efectivamente, el producto escalar entre dos vectores x_μ e y^μ transforma como

$$x'_\mu y'^\mu \;=\; x_\rho (\Lambda^{-1})^\rho{}_\mu \Lambda^\mu{}_\nu y^\nu \;=\; x_\rho \delta^\rho{}_\nu y^\nu \;=\; x_\rho y^\rho. \tag{5.14}$$

Igual que en el Capítulo 4, los operadores diferenciales $\partial_\mu \equiv \partial/\partial x^\mu$ y $\partial^\mu \equiv \partial/\partial x_\mu$ se comportan como vectores co- y contravariantes respectivamente,[5] como se puede averiguar fácilmente a través de la regla de la cadena. La divergencia (cuadrimensional) de un campo vectorial A^μ se puede escribir por lo tanto como el producto escalar entre los vectores ∂_μ y A^μ,

$$\partial_\mu A^\mu \;=\; \partial_t A^t + \vec{\nabla}\cdot\vec{A} \;=\; \partial_0 A^0 + \partial_1 A^1 + \partial_2 A^2 + \partial_3 A^3. \tag{5.15}$$

De la misma manera se puede ver el laplaciano cuadrimensional (a veces llamado el *d'alambertiano*) como la norma del operador ∂,

$$\Box \;\equiv\; \partial_\mu \partial^\mu \;=\; \partial_0^2 - \partial_1^2 - \partial_2^2 - \partial_3^2 \tag{5.16}$$

La invariancia del producto escalar (5.7) nos será de gran utilidad a la hora de formular una física covariante. En particular explica por qué las leyes de continuidad (leyes de conservación en teorías de campos) tienen la forma de una divergencia cuadrimensional (5.15) y en sistemas dinámicos el operador diferencial de segundo orden más común es el d'alambertiano (5.16).

La estructura del grupo de Lorentz es un poco más complicada que la del grupo ortogonal $O(N)$ del espacio euclídeo, debido a la presencia de la dirección temporal con el signo opuesto a las direcciones espaciales. Igual que en $\mathbb{R}^N$, sacando el determinante de (5.12), se ve que

$$\det \Lambda = \pm 1. \tag{5.17}$$

Esto divide el grupo de Lorentz en dos partes, según el signo del determinante. Las transformación con determinante -1 son reflexiones y no preservan la orientación de la base, mientras las transformaciones con determinante 1 son cambios de base que sí perservan la orientación. Sólo éstos últimos forman un subgrupo del grupo de Lorentz, puesto que la identidad pertenece a esta parte. Este subgrupo $\mathcal{L}_+$ se llama *el grupo propio de Lorentz*.

De (5.11) sacamos otra manera de dividir el grupo de Lorentz en dos partes: cogiendo la componente (00) de (5.11), vemos que

$$1 \;=\; (\Lambda^0{}_0)^2 - (\Lambda^1{}_0)^2 - (\Lambda^2{}_0)^2 - (\Lambda^3{}_0)^2. \tag{5.18}$$

Tenemos por lo tanto que la componente $\Lambda^0{}_0$ de una transformación de Lorentz tiene que satisfacer la condición de que $(\Lambda^0{}_0)^2 \geq 1$, o en otras palabras

$$\Lambda^0{}_0 \geq 1 \qquad \text{ó} \qquad \Lambda^0{}_0 \leq -1. \tag{5.19}$$

Las transformaciones con $\Lambda^0{}_0 \leq -1$ invierten la dirección temporal, mientras las que tienen $\Lambda^0{}_0 \geq 1$ la preservan. Sólo la parte con $\Lambda^0{}_0 \geq 1$ forma un subgrupo, llamado el *grupo ortocrono de Lorentz* $\mathcal{L}^\uparrow$.

Vemos por lo tanto que el grupo de Lorentz $\mathcal{L}$ se divide en 4 partes disconexas, caracterizadas por el signo del determinante y el signo de la componente $\Lambda^0{}_0$. De estas 4 partes, sólo la parte de las transformaciones propias y ortocronas $\mathcal{L}_+^\uparrow$ forma un subgrupo, puesto que contiene la identidad. Discutiremos ahora en más detalle las transformaciones de este subgrupo. Las transformaciones pertenecientes a las demás partes están relacionadas con las de $\mathcal{L}_+^\uparrow$ a través de reflexiones y/o cambios de orientación de la dirección temporal.

[5] Debido a los signos en la métrica, las componentes espaciales de ∂^μ y ∂_μ llevan el signo contrario. Para evitar confusión damos las componentes explícitamente: $\partial_\mu = (\partial_t,\ \vec{\nabla})$, mientras $\partial^\mu = \begin{pmatrix} \partial^t \\ -\vec{\nabla} \end{pmatrix}$.

El grupo de Lorentz propio y ortocrono $\mathcal{L}^{\uparrow}_{+}$ consiste en las transformaciones Λ que tienen determinante 1, preservan la dirección temporal y preservan la métrica $\eta_{\mu\nu}$ como en (5.11). Un ejemplo de una transformación son las rotaciónes en las tres dimensiones espaciales. Es fácil ver que por ejemplo una rotación en el plano xy

$$\Lambda^{\mu}{}_{\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\theta & \sin\theta & 0 \\ 0 & -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{5.20}$$

satisface todas las condiciones para pertenecer a $\mathcal{L}^{\uparrow}_{+}$. En otras palabras el grupo $SO(3)$ de rotaciones en $\mathbb{R}^3$ es un subgrupo de $\mathcal{L}^{\uparrow}_{+}$.[6]

Otro tipo de transformación que pertenece a $\mathcal{L}^{\uparrow}_{+}$ son las transformaciones especiales de Lorentz (3.19). Efectivamente, la matriz correspondiente

$$\Lambda^{\mu}{}_{\nu} = \begin{pmatrix} \gamma & -\gamma v & 0 & 0 \\ -\gamma v & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{5.21}$$

también preserva la métrica $\eta_{\mu\nu}$.

Es interesante darse cuenta de que podemos sustituir el parámetro v por

$$v = -\operatorname{tgh}\varphi, \tag{5.22}$$

de modo que $\gamma = 1/\sqrt{1-v^2} = \cosh\varphi$ y la transformación (5.21) se puede escribir como

$$\Lambda^{\mu}{}_{\nu} = \begin{pmatrix} \cosh\varphi & \sinh\varphi & 0 & 0 \\ \sinh\varphi & \cosh\varphi & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{5.23}$$

Esta forma de escribir un boost es remarcablemente parecida a una rotación espacial (5.20). La diferencia esencial es que en lugar de senos y cosenos aparecen sus equivalentes hiperbólicos. En cierto modo podemos pensar en un boost como una "rotación en el plano tx". Las funciones hiperbólicas aparecen por lo tanto debido al signo relativo entre la dirección temporal y la espacial.

Donde en una rotación (5.20) el parámetro de rotación θ corre de 0 a 2π, el parámetro φ de un boost puede tomar cualquier valor en $\mathbb{R}$, de $-\infty$ a ∞. Decimos por lo tanto que el grupo de Lorentz es un *grupo no-compacto*, porque el intervalo de algunos de sus parámetros no es compacto. Esta es la gran diferencia entre el grupo de Lorentz y el grupo $O(4)$ de las transformaciones ortogonales en $\mathbb{R}^4$. El grupo de Lorentz es el grupo de transformaciones ortogonales en un espacio 4-dimensional con una dirección temporal, lo que matemáticamente se anota respectivamente como $O(3,1)$ y $\mathbb{R}^{1,3}$. El subgrupo de transformaciones propias y ortocronas es por lo tanto el grupo $SO^{\uparrow}(3,1)$.

El grupo $\mathcal{L}^{\uparrow}_{+}$ (o equivalentemente $SO^{\uparrow}(3,1)$) consiste por lo tanto en las rotaciones espaciales y los boosts, que se interpretan como una rotación entre la dirección temporal y una espacial. $SO^{\uparrow}(3,1)$ tiene 6 parámetros independientes que tienen la siguiente interpretación física: un elemento de $SO^{\uparrow}(3,1)$ es una transformación entre dos sistemas de referencia arbitrarios en movimiento uniforme rectilíneo con la misma orientación de base y la misma orientación de la dirección temporal. Tres parámetros corresponden a los ángulos de Euler que relacionan la

[6] Y obviamente el grupo $O(3)$ es un subgrupo de $\mathcal{L}$ entero.

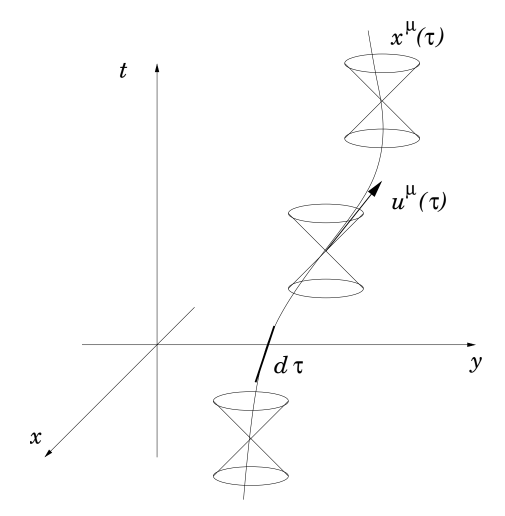


Figura 5.1: *La línea de universo de una partícula masiva es una curva temporal $x^\mu(\tau)$. El vector tangente a la curva, el cuadrivector velocidad, es un vector temporal. En cada punto, la curva entera está contenida dentro del cono de luz de este punto. La línea de universo se puede parametrizar de manera natural por el tiempo propio τ, el tiempo medido por el observador que se mueve con la partícula.*

orientación espacial de las dos bases. Otros dos parámetros corresponden dos ángulos indicando la dirección en que se mueve el segundo sistema de referencia con respecto al primero y el último parámetro corresponde a la velocidad relativa entre los dos sistemas.

Finalmente mencionamos que si no consideramos sólo las transfomaciones homogéneas (5.10) que mantienen el origen fijo, sino también permitimos transformación inhomogéneas

$$x'^\mu = \Lambda^\mu{}_\nu x^\nu + a^\mu, \tag{5.24}$$

es decir, incluimos también las traslaciones en el espacio y el tiempo, el grupo que forman estas transformaciones se llama el *grupo de Poincaré*. El grupo de Poincaré es el análogo relativista del grupo de Galilei y consta de 10 parámetros independientes: los 6 parámetros del grupo de Lorentz y 4 traslaciones en el espaciotiempo.

5.3. La dinámica relativista covariante

Repasaremos ahora la física relativista de la sección 3.3, pero en notación covariante, es decir sin referirnos a un sistema de referencia explícito.

La trayectoria de una partícula a través del espaciotiempo se llama la *línea de universo*. La línea de universo de una partícula con masa $m_0 \neq 0$ es una curva temporal, puesto que la velocidad de la partícula siempre es menor que la de la luz. Por lo tanto en cualquier punto de la trayectoria, toda la curva tiene que estar contenida dentro del cono de luz (pasado y futuro) en este punto (véase Figura 5.1).

Partículas sin masa se mueven a lo largo de curvas nulas. Esto es fácil de ver, puesto que la ecuación de una curva nula es $(x^0)^2 - (x^1)^2 - (x^2)^2 - (x^3)^2 = 0$, lo que es justo la ecuación (3.15) de un pulso de luz esférico expandiéndose desde el origen.

Se puede parametrizar la línea de universo de una partícula masiva utilizando el *tiempo propio* τ de la partícula: $x^\mu = x^\mu(\tau)$. El tiempo propio es el tiempo medido por un observador $\mathcal{O}$ que se mueve junto con la partícula y la relación entre un intervalo infinitesimal del tiempo propio $d\tau$ y

el tiempo dt medido por cualquier otro observador inercial $\mathcal{O}'$ viene dado por la fórmula (3.3) de la dilatación del tiempo:

$$d\tau = \sqrt{1-v^2}\, dt, \tag{5.25}$$

donde v es la velocidad de la partícula con respecto al observador $\mathcal{O}'$ en el intervalo $d\tau$. Para el observador $\mathcal{O}'$ el intervalo $d\tau$ corresponde a un intervalo entre dos puntos x^μ y $x^\mu + dx^\mu$ del espaciotiempo

$$d\tau^2 = dt^2 - dx^2 - dy^2 - dz^2 = \eta_{\mu\nu} dx^\mu dx^\nu. \tag{5.26}$$

Nótese que por lo tanto $d\tau$ es una cantidad invariante, dado que cualquier otro observador $\mathcal{O}''$ mediría el mismo intervalo. Esta propiedad es lo que hace que el tiempo propio sea particularmente útil para la parametrización de trayectorias. La longitud de la curva entera o de una parte finita se obtiene integrando $d\tau$ en un intervalo (in)finito.

El vector tangente a la curva es la *cuadrivelocidad* $u^\mu(\tau)$ y es por definición la derivada de la posición con respecto al tiempo propio

$$u^\mu(\tau) = \frac{dx^\mu}{d\tau}. \tag{5.27}$$

Para partículas masivas, u^μ es un vector temporal y de (5.26) se deriva fácilmente que entonces la norma de u^μ es (ejerc.)

$$\eta_{\mu\nu} u^\mu u^\nu = 1. \tag{5.28}$$

Para partículas sin masa, u^μ es un vector nulo.

Por construcción u^μ es un cuadrivector y transforma por lo tanto bien bajo el grupo de Lorentz. Una manifiestación de esto es el hecho de que la norma es un invariante, como nos indica (5.28). Físicamente se interpreta esta propiedad como que u^μ no es la velocidad newtoniana $\vec{v}$ de una partícula por el espacio, sino una velocidad que indica el movimiento por el espaciotiempo. De la propia definición de u^μ se ve que las tres componentes espaciales sí corresponden con las componentes de la velocidad newtoniana, $\vec{v} = d\vec{x}/dt$, aunque con el factor γ de corrección relativista, mientras la componente temporal es la velocidad de la luz, con el correspondiente factor relativista:

$$u^\mu \;=\; \frac{dx^\mu}{dt}\frac{dt}{d\tau} \;=\; \gamma \begin{pmatrix} 1 \\ \vec{v} \end{pmatrix}, \tag{5.29}$$

Se puede comprobar que actuando con un boost (5.21) sobre la cuadrivelocidad, $u'^\mu = \Lambda^\mu{}_\nu u^\nu$, se recupera la regla relativista de la suma de velocidades (3.22) (ejerc.).

De la misma manera se define la *aceleración* de una partícula como el cuadrivector[7]

$$\alpha^\mu(\tau) \;=\; \frac{du^\mu}{d\tau}. \tag{5.30}$$

Por su propia definición y la propiedad (5.28) vemos que en la mecánica relativista la velocidad u^μ y la aceleración α^μ siempre son ortogonales

$$u_\mu \alpha^\mu \;=\; \eta_{\mu\nu} u^\mu \frac{du^\nu}{d\tau} \;=\; \frac{1}{2}\frac{d}{d\tau}(\eta_{\mu\nu} u^\mu u^\nu) \;=\; 0. \tag{5.31}$$

La aceleración por lo tanto siempre es un vector espacial.

La aceleración relativista es bastante distinta a la aceleración newtoniana: en un sistema de referencia específico toma la forma

$$\alpha^\mu \;=\; \gamma \frac{du^\mu}{dt} \;=\;=\; \gamma \begin{pmatrix} d\gamma/dt \\ \vec{v} d\gamma/dt + \gamma \vec{a} \end{pmatrix} \;=\; \begin{pmatrix} \gamma^4 \vec{v}\cdot\vec{a} \\ \gamma^4 (\vec{v}\cdot\vec{a})\vec{v} + \gamma^2 \vec{a} \end{pmatrix}, \tag{5.32}$$

[7] A veces se oye que la relatividad especial no puede describir el movimiento acelerado y que para eso es preciso recurrir a la relatividad general. Esto es claramente erróneo: la única restricción que impone la relatividad especial es que se limita a describir la física desde el punto de vista de los observadores inerciales (i.e. observadores no-acelerados). Sin embargo, nada impide a un observador inercial estudiar la dinámica de una partícula acelerada.

donde $\vec{a} = d\vec{v}/dt$ es la aceleración newtoniana. Por lo tanto las componentes espaciales de α^μ no son simplemente (proporcionales a) la aceleración newtoniana, sino también tienen un término proporcional a $(\vec{v} \cdot \vec{a})\vec{v}$. Nótese además que el trivector $\vec{a}$ no es una cantidad invariante bajo transformaciones de Lorentz (y por lo tanto no es absoluta, $\vec{a}' \neq \vec{a}$), en contraste con lo que dijimos en la sección 2.2. En el fondo esto es de esperar, puesto que dos observadores distintos, que ven moverse la partícula con velocidades diferentes, deben medir aceleraciones diferentes, si queremos asegurar que en ningún sistema de referencia la velocidad de la partícula puede superar la velocidad de la luz.

El cuadrivector *momento* p^μ se define de manera análoga (pero, ojo, no igual) al momento newtoniano

$$p^\mu = m_0 u^\mu, \tag{5.33}$$

donde m_0 es la masa (de reposo) de la partícula. Dado que m_0 es un invariante y u^μ un cuadrivector, está claro que también p^μ lo es. De la definición se ve que las componentes espaciales de p^μ corresponden al momento newtoniano con un factor de corrección relativista $p^i = \gamma m_0 v^i$, mientras la componente temporal es la energía cinética E relativista,

$$p^\mu = \gamma \begin{pmatrix} m_0 \\ m_0\vec{v} \end{pmatrix} = \begin{pmatrix} E \\ \vec{p} \end{pmatrix}. \tag{5.34}$$

Bajo un boost (3.19), las componentes de p^μ transforman como (3.31). La norma de p^μ es un invariante y viene dada por

$$p_\mu p^\mu = m_0^2. \tag{5.35}$$

Nótese que esta expresión no es más que la formulación covariante de (3.32). La masa m_0 de una partícula es un invariante, simplemente porque es la norma del cuadrivector p^μ.

La versión relativista de la segunda ley de Newton viene dada por

$$f^\mu = \frac{dp^\mu}{d\tau}, \tag{5.36}$$

donde f^μ es el cuadrivector *fuerza*, cuyas componentes espaciales son proporcionales a la fuerza newtoniana $\vec{F}$ y la componente temporal $f^0 = \vec{v} \cdot \vec{F}$ al trabajo realizado por $\vec{F}$ por unidad de tiempo,

$$f^\mu = \gamma \begin{pmatrix} \vec{v} \cdot \vec{F} \\ \vec{F} \end{pmatrix} \tag{5.37}$$

Obsérvese que en términos de la aceleración cuadrimensional se puede escribir la segunda ley de Newton como

$$f^\mu = m_0 \alpha^\mu. \tag{5.38}$$

Sin embargo es importante darse cuenta de que, aunque formalmente análoga a $\vec{F} = m_0\vec{a}$, esta ecuación es tanto conceptualmente como físicamente muy distinta a su homóloga newtoniana.

De (5.36) podemos ver que mientras que no actúe ninguna fuerza f^μ sobre una partícula el cuadrivector p^μ se conserva. Obsérvese que por lo tanto la ley de la conservación de la energía y la ley de la conservación del momento, en la física newtoniana dos leyes independientes, en realidad corresponden a diferentes componentes de la misma ecuación cuadrimensional.

Finalmente, la definición del *momento angular* en la dinámica relativista es un poco más sutil. La generalización del momento angular tridimensional $\vec{\ell} = \vec{r} \times \vec{p}$ no es un cuadrivector, sino un tensor antisimétrico,[8]

$$L^{\mu\nu} = x^\mu p^\nu - x^\nu p^\mu, \tag{5.39}$$

[8]Esta definición del momento angular es general para cualquier número de dimensiones. El caso de un espacio tridimensional es especial, porque entonces un tensor antisimétrico es equivalente a un vector, a través del tensor de Levi-Civita: $\ell^i = \varepsilon^{ijk} x_j p_k$. Nótese que de este modo también recuperamos el conocido caso de que el momento angular en dos dimensiones viene dado por una sola componente, (es decir: es un escalar): $\ell = \varepsilon^{ij} x_i p_j$.

cuyas componentes $L^{ij} = \frac{1}{2}\varepsilon^{ijk}\ell_k$ nos dan el familiar momento angular tridimensional, mientras que las componentes $L^{0i} = tp^i - x^i E$ son una cantidad física de poca relevancia en la física newtoniana.

Derivando el momento angular con respecto al tiempo propio, obtenemos que

$$\begin{aligned} \frac{dL^{\mu\nu}}{d\tau} &= \frac{dx^\mu}{d\tau}p^\nu + x^\mu\frac{dp^\nu}{d\tau} - \frac{dx^\nu}{d\tau}p^\mu + x^\nu\frac{dp^\mu}{d\tau} \\ &= m_0(u^\mu u^\nu - u^\nu u^\mu) + x^\mu f^\nu - x^\nu f^\mu, \end{aligned} \tag{5.40}$$

donde en la segunda igualdad hemos utilizado la definición (5.33) del momento p^μ y la segunda ley de Newton (5.36). El primer término de esta igualdad es idénticamente cero por la antisimetría, mientras que en los últimos dos términos reconocemos el *momento de fuerza* relativista,

$$M^{\mu\nu} = x^\mu f^\nu - x^\nu f^\mu. \tag{5.41}$$

La ecuación (5.40) toma por lo tanto la forma

$$\frac{dL^{\mu\nu}}{d\tau} = M^{\mu\nu}, \tag{5.42}$$

lo que implica que el momento angular está conservado, mientras el momento de fuerza es cero.

Por otro lado, ya hemos dicho en las secciones 2.3 y 3.3 que las Leyes de Maxwell son covariantes bajo transformaciones de Lorentz y no necesitan por lo tanto una corrección relativista. Matemáticamente esto quiere decir que los objetos que aparecen en la teoría de Maxwell ya combinan de manera natural en cuadrivectores y tensores.

La densidad de carga ρ y la densidad de corriente $\vec{\jmath} = \rho\vec{v}$ forman las componentes de un cuadrivector j^μ y los potenciales ϕ y $\vec{A}$ combinan en el cuadrivector A^μ

$$j^\mu = \begin{pmatrix} \rho \\ \vec{\jmath} \end{pmatrix}, \qquad\qquad A^\mu = \begin{pmatrix} \phi \\ \vec{A} \end{pmatrix}. \tag{5.43}$$

Las transformaciones de j^μ y A^μ bajo un boost vienen dadas por (3.37) y (3.42).

Los potenciales ϕ y $\vec{A}$ dan lugar a los campos electromagnéticos $\vec{E}$ y $\vec{B}$ a través de las identidades (3.39), o en componentes

$$\begin{aligned} E_i &= -(\partial_t\vec{A})_i - (\vec{\nabla}\phi)_i = -\partial^0 A^i + \partial^i A^0, \\ B_i &= \epsilon_{ijk}(\vec{\nabla})_j(\vec{A})_k = -\epsilon_{ijk}\partial^j A^k = -\frac{1}{2}\epsilon_{ijk}(\partial^j A^k - \partial^k A^j). \end{aligned} \tag{5.44}$$

En otras palabras podemos definir un tensor antisimétrico $F^{\mu\nu}$ de rango $(2,0)$[9]

$$F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu, \tag{5.45}$$

cuyas componentes $(0i)$ son las componentes del campo eléctrico, $F^{0i} = -E_i$, y (ij) las del campo magnético, $F^{ij} = -\epsilon^{ijk}B_k$. El tensor $F^{\mu\nu}$ se llama el *tensor electromagnético*. Explícitamente tenemos (ejerc.)

$$F^{\mu\nu} = \begin{pmatrix} 0 & -E_1 & -E_2 & -E_3 \\ E_1 & 0 & -B_3 & B_2 \\ E_2 & B_3 & 0 & -B_1 \\ E_3 & -B_2 & B_1 & 0 \end{pmatrix}. \tag{5.46}$$

[9]El correspodiente tensor $F_{\mu\nu}$ de rango $(0,2)$ se define simplemente bajando los índices a través de la métrica: $F_{\mu\nu} = \eta_{\mu\rho}\eta_{\nu\lambda}F^{\rho\lambda}$. Ojo, el signo de algunas de las componentes de $F_{\mu\nu}$ será diferente que el de las componentes correspondientes de $F^{\mu\nu}$.

Dejamos como ejercicio averiguar que la transformación de $F^{\mu\nu}$ bajo las transformaciones de Lorentz $F'^{\mu\nu} = \Lambda^\mu{}_\rho \Lambda^\nu{}_\lambda F^{\rho\lambda}$ recupera las transformaciones (3.38) bajo el boost (5.21).

Con los vectores j^μ y A^μ y el tensor $F^{\mu\nu}$ tenemos todos los ingredientes para formular la teoría de Maxwell en su versión covariante. Las leyes homogéneas de Maxwell (3.33b) y (3.34a) se pueden escribir en función del tensor $F^{\mu\nu}$ como

$$\begin{aligned} (\vec{\nabla})_i B_i &= \frac{1}{2}\epsilon^{ijk}\partial_i F_{jk} = 0, \\ \epsilon_{ijk}(\vec{\nabla})_j E_k + \partial_t B_i &= -\frac{1}{2}\epsilon_{ijk}\Big(\partial^j F^{k0} + \partial^k F^{0j} + \partial^0 F^{jk}\Big) = 0. \end{aligned} \tag{5.47}$$

Aunque a primera vista estas dos ecuaciones parecen muy diferentes, se trata en realidad de la componente temporal y las componentes espaciales de la misma ecuación cuadrimensional. Esto se ve al darse cuenta de que la relación entre el tensor de Levi-Civita tri- y cuadrimensional viene dada por $\epsilon^{0ijk} = \epsilon^{ijk}$. Las dos ecuaciones (5.47) se pueden por lo tanto resumir en

$$\epsilon^{\mu\nu\rho\lambda}\partial_\nu F_{\rho\lambda} = 0, \tag{5.48}$$

o equivalentemente

$$\partial_\nu F_{\rho\lambda} + \partial_\lambda F_{\nu\rho} + \partial_\rho F_{\lambda\nu} = 0. \tag{5.49}$$

La ecuación (5.48) ó (5.49) se llama la *identidad de Bianchi* y la solución de esta ecuación viene dada por (5.45) (ejerc.). En otras palabras, la identidad de Bianchi (o equivalentemente las ecuaciones homogéneas de Maxwell) nos dice que se puede escribir $F_{\mu\nu}$ como la derivada antisimetrizada de un potencial A_μ, tal como dice (5.45).

En realidad hay muchos potenciales que dan el mismo $F_{\mu\nu}$: dado un potencial A_μ que genera una $F_{\mu\nu}$ a través de (5.45), el potencial A'_μ relacionado con A_μ a través de

$$A'_\mu = A_\mu + \partial_\mu \Lambda, \tag{5.50}$$

con Λ una función arbitraria, también genera la misma expresión para $F_{\mu\nu}$. Esta ambigüedad de los potenciales A^μ es obviamente la traducción a lenguaje cuadrimensional de las transformaciones gauge (1.50), y nos está diciendo es que los A^μ no son más que unos campos auxiliares, mientras que la física real está en el tensor electromagnético $F_{\mu\nu}$. Hasta cierto punto el rol de la invariancia gauge en la física de particulas es comparable con el rol de la invariancia bajo transformaciones de coordenadas en la teoría de la relatividad. En las matemáticas se han encontrado profundas relaciones entre estos dos principios en la teoría de fibrados.

Las ecuaciones de Maxwell que nos quedan, las inhomogéneas (3.33a) y (3.34b) también se pueden escribir en función de objetos covariantes. Un pequeño cálculo muestra que estas dos ecuaciones dan respectivamente (ejerc.)

$$\begin{aligned} \partial_j F^{j0} &= j^0 \\ \partial_j F^{ji} + \partial_0 F^{0i} &= j^i. \end{aligned} \tag{5.51}$$

tambien aquí resulta que las dos ecuaciones corresponden a la componente temporal y las componentes espaciales de la misma ecuación cuadrimensional. Para ver esto, no hace falta mas que añadir a la primera ecuación el término $\partial_0 F^{00}$, que es identicamente cero por la antisimetría de $F^{\mu\nu}$. Por lo tanto las ecuaciones inhomogéneas se combinan en

$$\partial_\mu F^{\mu\nu} = j^\nu. \tag{5.52}$$

Vemos por lo tanto que hemos resumido las cuatro ecuaciones de Maxwell (3.33a) y (3.34b) en dos, siendo (5.48) y (5.52). Por su estructura de índices está claro que las dos son ecuaciones vectoriales[10] y transforman como un cuadrivector bajo las transformaciones de Lorentz (ejerc.), como ya habíamos visto en el Capítulo 3.

[10]La identidad de Bianchi es un tensor de rango (0,3) en su formulación (5.49).

De la ecuación inhomogénea de Maxwell podemos derivar directamente otra relación importante, tomando la divergencia de (5.52), es decir actuando con la derivada ∂_ν. El lado izquierdo entonces es cero, por la combinación de la simetría de las derivadas y la antisimetría de $F^{\mu\nu}$, de modo que obtenemos

$$\partial_\mu j^\mu = 0. \tag{5.53}$$

Esto es la ley de conservación de carga, que hemos visto en notación tridimensional en (3.36). Nótese que esta ecuación tiene la estructura del producto escalar entre entre ∂_μ y j^μ y por lo tanto es invariante bajo las transformaciones de Lorentz. La conservación de la carga es por lo tanto un hecho que no depende de un observador, sino que es válido para todos los observadores (inerciales).

También podemos recuperar las ecuaciones inhomogéneas de ondas (3.40) que satisfacen los potenciales, sustituyendo la expresión (5.45) en (5.52). Asumiendo que los potenciales satisfacen el gauge de Lorenz (3.41)

$$\partial_\mu A^\mu = 0, \tag{5.54}$$

vemos que (5.52) se reduce a la versión covariante de las ecuaciones (3.40)

$$\partial_\mu \partial^\mu A^\nu = j^\nu. \tag{5.55}$$

Finalmente es útil observar que se puede derivar toda la dimámica mencionada en esta sección a través del formalismo lagrangiano de la acción de la partícula cargada y para el campo electromagnético (ejerc.)

$$\begin{aligned} S_{\text{partícula}} &= -\frac{1}{2}m_0 \int d\tau\, \eta_{\mu\nu}\dot{x}^\mu \dot{x}^\nu \;-\; q\int d\tau\, \dot{x}_\mu A^\mu(x(\tau)), \\ S_{\text{Maxwell}} &= -\frac{1}{4}\int d^4x\, F_{\mu\nu}(x)F^{\mu\nu}(x) \;-\; \int d^4x\, j_\mu A^\mu. \end{aligned} \tag{5.56}$$

Es muy instructivo averiguar que estas acciones son precisamente las acciones (1.60) y (1.63) en lenguaje convariante (aunque la primera con las necesarias correcciones relativistas para que sea invariante Lorentz) y por lo tanto toda la física comentada en la sección 1.5 (variables dinámicas, invariancia gauge, etc) aplica de la misma manera a estas acciones (ejerc.).

5.4. La necesidad de la relatividad general

Ya en 1907, sólo dos años después de la publicación de la relatividad especial, Einstein se dio cuenta de que la teoría de la gravedad newtoniana y la relatividad especial son mutuamente incompatibles (salvo en el caso de un campo gravitatorio constante y estático). Hay varias maneras, matemáticas y físicas, de ver esto.

Matemáticamente, se ve porque la gravedad newtoniana no es invariante bajo el grupo de Lorentz. Según Newton, una partícula en un campo gravitatorio está sometida a una aceleración

$$\frac{d^2\vec{x}}{dt^2} = -\vec{\nabla}\Phi, \tag{5.57}$$

causada por la fuerza gravitatoria, cuyo potencial Φ está relacionado con la densidad de materia ρ_m en el universo a través de la *ecuación de Poisson*

$$\Delta\Phi = 4\pi G_N \rho_M, \tag{5.58}$$

donde G_N es la constante de Newton. Obviamente, ni (5.57), ni (5.58) transforman bien bajo una transformación de Lorentz.

Primero, tanto el lado izquierdo como el derecho de (5.57) son vectores tridimensionales y no cuadrimensionales. Además, la aceleración está definida como la segunda derivada de la posición con respecto al tiempo, pero no está claro con respecto al tiempo de qué observador. Para Newton esto no era ningún problema, puesto que para él existía un solo tiempo absoluto, igual para todos los observadores. Sin embargo de la teoría de la relatividad especial sabemos que cada observador tiene su tiempo propio particular.

Se podría intentar remediar estos problemas, derivando con respecto al tiempo propio e intentando convertir (5.57) en una ecuación covariante, como acabamos de hacer con la mecánica newtoniana. Pero un problema más gordo nos supone la otra parte de la teoría de la gravedad newtoniana, la ecuación (5.58). Aquí aparece el laplaciano $\Delta\Phi$, en lugar del operador invariante, el d'alambertiano $\Box\Phi$. Para apreciar las dificultades que supone este operador en lugar del otro, pasaremos a los argumentos físicos de la incompatibilidad.

La ecuación (5.58) dice que el potencial gravitatorio Φ en un punto x está determinado por la distribución de materia ρ_M en el universo. Si por lo tanto un observador cambia la distribución de materia en cierto punto, el efecto en el potencial gravitatorio se nota inmediatamente en todo el universo. En otras palabras, la fuerza gravitatoria se propaga en la teoría de Newton con una velocidad infinita.[11]

No sólo velocidades mayores que la velocidad de la luz son problemáticas en la relatividad especial, también surge una pregunta básica: Si el cambio de la distribución de materia en un punto y el cambio del potencial en otro punto son simultáneos, ya que el efecto se nota de manera inmediata, ¿son simultáneos para qué observador? Hemos visto en el Capítulo 3 que la simultaneidad de sucesos es algo que no está bien definido, sino que depende del observador. Si la gravedad newtoniana tiene que recurrir a un observador especial, para el cual las fórmulas (5.57) y (5.58) son válidas, viola el Principio de la Relatividad.

El verdadero problema de la ecuación de Poisson (5.58) es que es válida para el caso estático, cuando la distribución de materia no varía. En este sentido, (5.58) es el análogo gravitacional de la ley de Gauss $\vec{\nabla}\cdot\vec{E} = \rho_e$ en electrostática, cuando la distibución de cargas eléctricas es constante en el tiempo. En este caso podemos elegir el gauge de Coulomb, $\vec{\nabla}\cdot\vec{A} = 0$, de modo que se puede escribir la ecuación de Gauss completamente en términos del potencial eléctrico ϕ,

$$\Delta\phi \;=\; -\rho_e. \tag{5.59}$$

Sin embargo, esta descripción es insuficiente en el caso de fuentes no-estáticas: los campos magnéticos $\vec{B}$ que aparecen en ese caso nos obligan a introducir un potencial vectorial $\vec{A}$, cuya dinámica, igual que la de ϕ, está descrita por una ecuación de onda. Las soluciones de ésta son los conocidos potenciales retardados, que describen la propagación con velocidad finita de las perturbaciones, compatibles con la causalidad relativista. En otras palabras, covariancia Lorentz e invariancia gauge nos dicen que la generalización de la ley de Gauss (5.59) al caso no-estático es directamente la teoría de Maxwell entera.

En 1907 Einstein se vio confrontado con el equivalente gravitacional: encontrar la teoría dinámica de la gravedad, que sea compatible con las exigencias básicas de la relatividad especial (en particular, la propagación a velocidad finita) y cuyo caso estático sea la ecuación de Poisson (5.58). Si el paso de electrostática a la teoría de Maxwell es grande, el de la gravedad newtoniana a la relatividad general lo es más aún y a Einstein le costó mucho remediar este problema. Aunque ya se dio cuenta en 1907 de la incompatibilidad de ambas teorías y de la solución, el Principio de Equivalencia, tardó hasta 1911 en llegar a una primera formulación matemática y otros 2 años más, hasta 1913, en el Principio de Covariancia. Y no fue hasta 1915, diez años después de la relatividad especial, cuando vino con una versión definitiva de la relatividad general.

El problema que surge al intentar incorporar la gravedad es que tenemos que tratar con obser-

[11] Nótese que esto no pasaría si el operador diferencial fuera un d'alambertiano: las soluciones en este caso son funciones del tipo $f(t \pm x)$, los conocidos potenciales retardados y avanzados.

vadores acelerados y por lo tanto no-inerciales (o por lo menos tenemos que tratar con observadores que no pertenecen a la misma clase inercial), los cuales la relatividad especial simplemente no considera. La solución al problema consiste básicamente en un viejo truco de la mecánica clásica: la introducción de fuerzas ficticias.

En la mecánica newtoniana un sistema inercial está definido como un sistema en el cual la Segunda Ley de Newton, $\vec{F} = m\vec{a}$, es válida, es decir un sistema en el que cada aceleración $\vec{a}$ de una partícula está debida a una fuerza $\vec{F}$ que actúa sobre la partícula. En sistemas no-inerciales por otro lado, las partículas libres no seguirán trayectorias rectas con respecto a este sistema, debido a la aceleración de este mismo. Sin embargo, por razones prácticas un observador no-inercial puede auto-engañarse y creerse inercial, siempre y cuando asuma la existencia de fuerzas ficticias $\vec{f}$ que se crean responsables de las aceleraciones no-inerciales $\vec{a}_{\text{n.i.}}$ observadas, de modo que la Segunda Ley de Newton vuelva a ser válida otra vez:

$$\vec{F} + \vec{f} = m(\vec{a} + \vec{a}_{\text{n.i.}}). \tag{5.60}$$

Ejemplos conocidos de la mecánica newtoniana son la fuerza de Coriolis y la fuerza centrífuga.

Lo que hizo Einstein en realidad fue darse cuenta de que partículas en caída libre son (localmente) equivalentes a partículas libres en un sistema inercial, sobre las cuales no actúa ninguna fuerza y que por lo tanto podía considerar la gravedad como una fuerza ficticia. Esto es un paso altamente no-trivial y sólo es posible gracias a que todos los objetos caen con la misma velocidad, independientemente de su masa. No es ninguna coincidencia que en (5.60) la masa (gravitatoria) delante de la aceleración gravitatoria $\vec{a}$ sea la misma masa que la masa (inercial) delante de la aceleración no-inercial $\vec{a}_{\text{n.i.}}$. Esto se llama el Principio de Equivalencia y lo discutiremos en más detalle en el Capítulo 9.

Una de las peculiaridades del Principio de Equivalencia es que es solamente válido localmente, es decir a escalas pequeñas, debido a las inhomogeneidades de un campo gravitatorio genérico. Dos observadores en caída libre, uno encima del polo norte y otro encima del ecuador, se verán mutuamente acelerados, ya que siguen trayectorias ortogonales hacia el centro del la Tierra. Aunque cada uno se puede considerar a sí mismo como un observador inercial, no existe ningún sistema que contenga a los dos observadores a la vez como inerciales. El sistema inercial que asume cada observador tiene una extensión finita y pequeña.

Esta observación tiene una consecuencia profunda: implica que se puede considerar un espaciotiempo con gravedad como un conjunto de pequeños trozos de espaciotiempo plano (Minkowski), pero no hay un sistema inercial global. La unión de todos los trocitos de espacios planos "pegados y cosidos" forma un espacio que globalmente no es plano, sino que tiene curvatura. En realidad, lo que consiguió Einstein con la relatividad general fue interpretar la gravedad como una fuerza ficticia, en realidad debida a la geometría curva del espaciotiempo. Las trayectorias de partículas aceleradas por la gravedad no son más que el análogo de líneas rectas (trayectorias de partículas libres) en un espacio curvo.

Matemáticamente, una cosa que globalmente es un espacio con curvatura, pero localmente tiene el aspecto de un espacio plano, se llama una *variedad* y la herramienta matemática que permite hacer álgebra y análisis en un espacio curvo es la *geometría diferencial*, una generalización de la geometría plana de Euclides. De la discusión arriba está claro que la geometría diferencial es una herramienta indispensable para estudiar la Relatividad General, de modo que en la Parte II la estudiaremos en más detalle.

Parte II

Geometría Diferencial

Capítulo 6

Variedades y cambios de coordenadas generales

No hay Camino Real hacia la geometría.
(Euclides, cuando Ptolomeo I le preguntó si no había
una manera más sencilla de aprender matemáticas)

En la sección 5.4 hemos visto que el intento de incluir la gravedad en el formalismo de la relatividad especial sugiere que el espaciotiempo no es $\mathbb{R}^{1,3}$, sino algo que localmente parece plano, aunque globalmente no necesariamente lo sea. La geometría de estas variedades curvas es una generalización de la geometría euclídea y en la Parte II de este libro introduciremos las herramientas matemáticas necesarias para trabajar en variedades curvas, en particular las propiedades algebráicas y geométricas que resultarán imprescindibles para la descripción relativista de la gravedad.

6.1. Breve historia de la geometría no-euclídea

Los Elementos de Euclides (ca. 325 - ca. 265 a. C.) es sin duda uno de los libros más influentes de la historia de la ciencia.[1] Escrito en Alejandría (Egipto) alrededor del año 300 a. C., la obra consiste en realidad en 13 libros que tratan la mayoría de las matemáticas conocidas entonces, pero los Elementos son realmente conocidos por su estudio de la geometría plana y la geometría espacial. Durante más de 2000 años los Elementos han sido el texto estándar por excelencia sobre la geometría de espacios planos, tanto que esta última se suele denominar también *geometría euclídea*.

La causa de la admiración que los matemáticos siempre han sentido hacia los *Elementos* es su estricta metodología: partiendo solamente de definiciones y postulados, Euclides logra demostrar cerca de 500 proposiciones (teoremas), utilizando sólo las definiciones, postulados y proposiciones anteriormente demostrados. Este rigor, que respira toda la obra, ha influenciado a generaciones de científicos y durante siglos el estudio de los *Elementos* formaba parte de cualquier carrera universitaria.

De los cinco postulados que usa Euclides para construir su geometría, los primeros cuatro son afirmaciones bastante directas sobre puntos, rectas, círculos y angulos rectos, pero el quinto

[1] Desde su primera edición imprenta en 1482 en Venecia, han sido publicadas por lo menos 1000 ediciones. Sólo de la Biblia se han publicado más ediciones.

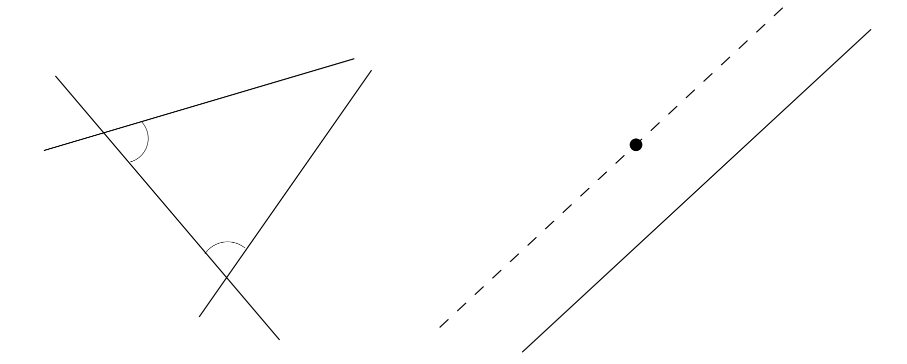


Figura 6.1: *El Quinto Postulado de Euclides, en la formulación original de Euclides (izquierda) y la formulación de Playfair (derecha). La aceptación o negación de Quinto Postulado lleva a geometrías distintas, pero todas consistentes.*

es muy diferente y ha obtenido un lugar particularmente importante en la historia de las matemáticas. En formulación de Euclides afirma que

> Y que si una recta al incidir sobre dos rectas hace los ángulos internos del mismo lado menores que dos rectos, las dos rectas prolongadas indefinidamente se encontrarán en el lado en el que están los [ángulos] menores que dos rectos,

aunque es más conocida la formulación del matemático escocés John Playfair (1748 - 1819) de 1795, por lo que a veces recibe el nombre del *Postulado de los Paralelos*,

> Por un punto exterior a una recta dada sólo cabe trazar una paralela.

Su carácter distinto siempre le ha dado un estatus especial y aunque su veracidad parece obvia a primera vista (véase Figura 6.1), se ha llegado a preguntar si es necesario incluirlo como un postulado.[2] Debido al hecho de que su formulación parece más la de un teorema que de un postulado, generaciones de matemáticos, como el italiano Giovanni Saccheri (1667 -1733) y el suizo Johann Lambert (1728 - 1777), han intentado derivar el Quinto Postulado de los demás, lo que lo hubiera convertido en una proposición. Aunque ninguno de estos intentos logró su objetivo,[3] dieron lugar a resultados importantes de lo que más tarde iba a llamarse la *geometría no-euclídea*.

Alrededor de 1830 los matemáticos húngaro János Bolyai (1802 - 1860) y ruso Nikolai Lobachevsky (1792 - 1856) desarrollaron independientemente una *geometría hiperbólica*, en que la se reemplaza el Quinto Postulado por la asunción de que por un punto exterior a una recta pasen infinitas rectas. Aunque la visualización de tal geometría desafía la visualización humana, el logro de Bolyai y Lobachevsky fue darse cuenta de que esa asunción no lleva (necesariamente) a inconsistencias. Cuando el padre de Bolyai, íntimo amigo del matemático alemán Carl Friedrich Gauss (1777 - 1855), le comentó a éste las ideas de su hijo, Gauss admitió que él también había obtenido resultados parecidos, pero que no se atrevía a publicarlos porque temía que pudiera haber perjudicado su carrera.

Las preguntas que se planteaban entonces a raíz de los descubrimientos de Bolyai y Lobachevsky eran primero si los nuevos sistemas eran consistentes, en el sentido que no llevaban a

[2]Hasta Euclides mismo parece haberse dado cuenta de lo extraño que es este postulado, ya que no lo utiliza en las primeras 28 proposiciones que demuestra, como si quisiera aplazar su uso lo máximo posible.

[3]Visto a posteriori, básicamente por aferrarse más a la intuición euclídea que a consistencia interna. En cierto momento Saccheri creía haber demostrado la necesidad del Quinto Postulado, al encontrar que no asumirlo llevaba a resultados que eran *"repugnantes a la naturaleza de la recta"*.

contradicciones lógicas, y segundo si había alguna razón para considerar la geometría euclídea como más fundamental que las demás posibilidades.

Una respuesta, aunque parcial, a la primera pregunta fue proporcionada por el matemático alemán Felix Klein (1849 - 1925) en 1872. Klein demostró que se puede proyectar el plano de Lobachevsky en una región finita del plano euclídeo y que existe una relación una a una entre las afirmaciones de la geometría hiperbólica y las afirmaciones de la geometría euclídea. Con esto no quedaba demostrada que la geometría hiperbólica estuviera libre de contradicciones, pero su consistencia interna ahora quedaba relacionada con la de la euclídea: encontrar una contradicción en la geometría hiperbólica, se traduciría a una contradicción en la geometría euclídea, lo que parecía mucho más improbable. Más tarde, el matemático alemán David Hilbert (1862 - 1943) relacionó la consistencia interna de la geometría euclídea con la consistencia interna de la aritmética.

La respuesta a la segunda pregunta era más sútil. Poco a poco los matemáticos empezaron a darse cuenta de que la geometría euclídea y la hiperbólica no eran mutuamente excluyentes, sino que formaban dos construcciones matemáticas distintas, basadas en sistemas axiomáticos distintos. Una vez aceptada la posibilidad de tener diferentes geometrías consistentes, dependiendo de si se acepta o no el Quinto Postulado, se reconoció la geometría esférica, conocida de la cartografía desde la Antigüedad, como otra variante más, que se caracteriza por asumir que no existe ninguna recta que pase por un punto y sea paralela a una recta dada.[4]

Por lo tanto, la pregunta sobre la veracidad del Quinto Postulado o de una de sus variantes resulta irrelevante: los axiomas no son verdad o falso, son afirmaciones arbitrarias sobre las cuales se construyen distintos sistemas lógicos. Por otro lado, la pregunta de cuál de las (por lo menos) tres geometrías distintas es "más realista", se ha convertido en una pregunta física, ya que hace referencia al mundo en que vivimos. La matemática debe estudiar las propiedades de las distintas geometrías y la física debe determinar cual de ellas corresponde con la Naturaleza.

En los años 1850, Gauss y su estudiante, el matemático alemán Bernhard Riemann (1826 - 1866), desarrollaron un enfoque distinto a la geometría no-euclídea, que en vez de axiomático, era analítico. Gauss ya se había dado cuenta de que el concepto de métrica resulta muy útil para describir las propiedades geométricas de superficies bidimensionales embebidas en $\mathbb{R}^3$. La idea es que una hormiguita bidimensional puede llegar a descubrir que está en una superficie curva, sin saber de la existencia de una dirección ortogonal a la superficie. En una esfera (un hiperboloide) la suma de los ángulos de un triangulo es mayor (menor) que π y la proporción entre la circunferencia y el radio de un círculo es menor (mayor) que 2π, de modo que la hormiguita puede distinguir estos casos a través de las mediciones apropiadas.

Gauss había descubierto por lo tanto que la esfera y el hiperboloide, y en general cada superficie, tienen propiedades geométricas intrísecas que las distingue del plano y entre ellos. Riemann generalizó esta idea de geometría intrínseca a espacios de dimensiones arbitrarias y definió el concepto de curvatura del espacio, codificado en el tensor de Riemann, a través de la idea de transporte paralelo. El enfoque de Riemann es una generalización de las ideas de Gauss, en el sentido de que las propiedades geométricas intrénsicas de un espacio se pueden deducir sin considerar que el espacio esté embebido en un espacio más grande, sino sólo por mediciones internas. Tan influyente fue el trabajo de Riemann que hoy en día la geometría de los espacios curvos se denomina *geometría riemanniana*.

Alrededor de 1907, cuando Einstein estaba intentando incorporar la gravedad en su teoría de la relatividad, la geometría riemannniana no formaba parte de la formación estándar de un físico. Era por eso que el matemático suizo Marcel Grossmann (1878 - 1936), amigo y compañero de clase de Einstein en Zurich, llegó a jugar un papel importante en el desarrollo de la relatividad general, indicando a Einstein la importancia de la teoría de Riemann e instruyéndole en

[4]La única sutileza era ampliar la definición de una recta a una geodésica, la curva más corta entre dos puntos, en caso de una esfera un círculo máximo (véase la sección 8.8).

ello. La única diferencia entre la geometría diferencial que necesitaba Einstein en su teoría de la relatividad general y la desarrollada por Riemann es la signatura de la métrica: donde Riemann sólo consideraba espacios donde todas las direcciones son espaciales, Einstein se vió obligado a trabajar con espaciotiempos, que tienen una coordenada temporal, puesto que Minkowski ya había formulado la relatividad especial en términos de un espacio plano lorentziano.

En este capítulo definiremos el concepto de variedad y estudiaremos sus propiedades algebraicas, mientras que en los siguientes dos capítulos introduciremos la teoría de Riemann sobre curvatura. En principio trabajaremos con variedades riemannianas (con direcciones espaciales), pero en realidad todos los resultados relevantes impotantes son directamente aplicables a espacios lorentzianos.

6.2. Repaso de coordenadas curvilíneas en $\mathbb{R}^N$

Antes de empezar a estudiar cambios de coordenadas en espacios arbitrarios, repasaremos primero algo más familiar: la descripción de $\mathbb{R}^N$ en coordenadas curvilíneas. Ejemplos conocidos en $\mathbb{R}^3$ son las coordenadas esféricas

$$x = r\sin\theta\cos\varphi, \qquad y = r\sin\theta\sin\varphi, \qquad z = r\cos\theta. \tag{6.1}$$

o las coordenadas cilíndricas

$$x = \rho\cos\varphi, \qquad y = \rho\sin\varphi, \qquad z = z. \tag{6.2}$$

Es importante darse cuenta de que, aunque en coordenadas curvilíneas las expresiones de la métrica y el producto escalar pueden tener un aspecto muy diferente a en coordenadas cartesianas, sólo estamos aplicando un cambio de coordenadas y por lo tanto todas las propiedades geométricas del espacio aun siguen siendo las de $\mathbb{R}^N$.

Consideramos ahora N funciones invertibles $y^\mu(x^i)$ de las coordenadas cartesianas x^i. El hecho de que las funciones $y^\mu(x^i)$ sean invertibles implica que a cada punto de $\mathbb{R}^N$ se le puede asignar un único valor de y^μ y que cada valor de y^μ corresponde a un solo punto de $\mathbb{R}^N$. Las funciones y^μ sirven por lo tanto igual de bien como coordenadas sobre $\mathbb{R}^N$ que las coordenadas cartesianas x^i.[5] La transformación

$$y^\mu = y^\mu(x^i) \tag{6.3}$$

se llama una *transformación general de coordenadas*. La gran diferencia entre una transformación (6.3) y la transformación (4.5) es que en (6.3) las reglas de transformación varían de punto un punto, mientras en (4.5) son las mismas para todo $\mathbb{R}^N$. El cambio de coordenadas (4.5) se llama una *transformación global* y (6.3) una *transformación local*.

Igual que en el sistema cartesiano, podemos definir una base $\{|e_\mu\rangle\}$ para las coordenadas curvilíneas. Un vector contravariante $|V\rangle$, que en una base cartesiana $\{|e_i\rangle\}$ se descompone como $|V\rangle = V^i|e_i\rangle$, tendrá en coordenadas curvilíneas la forma $|V\rangle = V^\mu|e_\mu\rangle$, donde los $|e_\mu\rangle$ son los vectores de base (no necesariamente ortogonales o normalizados). Descomponiendo los vectores de base $|e_\mu\rangle$ en la base $\{|e_i\rangle\}$ obtenemos que

$$|e_\mu\rangle = \frac{\partial |r\rangle}{\partial y^\mu} = \frac{\partial x^i}{\partial y^\mu}\,|e_i\rangle. \tag{6.4}$$

La primera igualdad identifica los vectores de base $|e_\mu\rangle$ como los tangentes a las curvas de coordenadas y^μ, mientras que la segunda igualdad relaciona los vectores de una base con la otra. Nótese que la matriz de transformación es $M^i{}_\mu = \partial x^i/\partial y^\mu$.

[5] Para evitar confusión utilizamos índices latinos $(i, j, k, ...)$ para las coordenadas cartesianas e índices griegos $(\mu, \nu, \rho, ...)$ para las coordenadas curvilíneas.

Por un argumento idéntico al de la sección 4, es fácil de averiguar que entonces las componentes de $|V\rangle$ en las distintas coordenadas están relacionadas como

$$V^\mu = \frac{\partial y^\mu}{\partial x^i}\, V^i. \tag{6.5}$$

Obsérvese que las reglas de transformación entre la base cartesiana y la curvilínea dependen del punto p en $\mathbb{R}^N$ en que se haga la transformación, debido al carácter local de la transformación (6.3). Esto implica que, en contraste con el sistema cartesiano, una dirección $|e_\mu\rangle$ varía de punto en punto. Por ejemplo, la dirección radial $|e_r\rangle$ en coordenadas polares es distinta en puntos distintos. Esto tendrá consecuencias importantes, como veremos más adelante.

También los vectores covariantes se pueden descomponer en una base de coordenadas curvilíneas $\langle W| = W_\mu \langle e^\mu|$, donde los vectores de la base dual actúan sobre los vectores de la base $\{|e_\mu\rangle\}$ como

$$\langle e^\mu | e_\nu\rangle = \delta^\mu{}_\nu \tag{6.6}$$

Para mantener esta relación bajo cambios de coordenadas es preciso que los vectores de la base dual y las componentes de los vectores covariantes transformen como

$$\langle e^\mu| = \frac{\partial y^\mu}{\partial x^i}\langle e^i|, \qquad\qquad V_\mu = \frac{\partial x^i}{\partial y^\mu}\, V_i. \tag{6.7}$$

A través de las reglas de transformación (6.4) y (6.7) podemos encontrar una expresión para la métrica en coordenadas curvilíneas. Aplicando los cambios de coordenadas a (4.21) obtenemos que

$$\frac{\partial y^\mu}{\partial x^i}|e_\mu\rangle = \delta_{ij}\frac{\partial x^j}{\partial y^\nu}\langle e^\nu|, \tag{6.8}$$

y multiplicando esta ecuación con $\partial x^i/\partial y^\rho$ encontramos que

$$|e_\rho\rangle = \delta_{ij}\frac{\partial x^i}{\partial y^\rho}\frac{\partial x^j}{\partial y^\nu}\langle e^\nu|. \tag{6.9}$$

Comparando esta relación con (4.21) sugiere que interpretemos los factores delante de $\langle e^\rho|$ como la métrica de $\mathbb{R}^N$ en coordenadas curvilíneas:

$$g_{\mu\nu} = \frac{\partial x^i}{\partial y^\mu}\frac{\partial x^j}{\partial y^\nu}\delta_{ij}. \tag{6.10}$$

Obsérvese que esta relación también nos da directamente las reglas de transformación entre la forma de la métrica en coordenadas cartesianas y curvilíneas. Es un ejercicio instructivo comprobar que una manera alternativa de encontrar la forma de la métrica en coordenadas curvilíneas es la equivalente de (4.29), es decir tomando el producto escalar entre los vectores $|e_\mu\rangle$:

$$g_{\mu\nu} = |e_\mu\rangle \cdot |e_\nu\rangle. \tag{6.11}$$

Por ejemplo, un pequeño ejercicio muestra que la métrica de $\mathbb{R}^3$ en coordenadas cilíndricas y esféricas viene dada por

$$\begin{aligned} ds^2_{\text{cil}} &= d\rho^2 + \rho^2 d\varphi^2 + dz^2, \\ ds^2_{\text{esf}} &= dr^2 + r^2 d\theta^2 + r^2\sin^2\theta d\varphi^2. \end{aligned} \tag{6.12}$$

Es importante darse cuenta de que, aunque la expresión de la métrica en coordenadas curvilíneas es muy distinta a la de en coordenadas cartesianas, sólo hemos hecho un cambio de coordenadas. El espacio que describe $g_{\mu\nu}$ es todavia $\mathbb{R}^N$ con las mismas propiedades geométricas que en coordenadas cartesianas. Un cambio de coordenadas sólo es una manera distinta de etiquetar los puntos de un espacio y nunca puede cambiar las propiedades geométricas de éste.

La métrica $g_{\mu\nu}$ juega el mismo papel que la métrica euclídea cartesiana δ_{ij}. En particular, la métrica y su inversa suben y bajan los índices de vectores co- y contravariantes

$$V_\mu = g_{\mu\nu} V^\nu, \qquad\qquad V^\mu = g^{\mu\nu} V_\nu, \tag{6.13}$$

y el producto escalar de dos vectores en coordenadas curvilíneas está definido como

$$\langle V|W\rangle \;=\; V_\mu W^\mu \;=\; g_{\mu\nu} V^\mu W^\nu. \tag{6.14}$$

Las operaciones diferenciales son un poco más sutiles en coordenadas curvilíneas que en coordenadas cartesianas. Donde en coordenadas cartesianas la derivada de un campo vectorial es simplemente la derivada de cada componente

$$\frac{\partial |V\rangle}{\partial x^i} \;=\; \frac{\partial V^j}{\partial x^i}\, |e_j\rangle, \tag{6.15}$$

en coordenadas curvilíneas el operador diferencial también actúa sobre los vectores de base $|e_\mu\rangle$, puesto que éstos cambian de punto en punto. Por lo tanto la derivada de un vector $\vec{V}$ está dada por

$$\partial_\mu |V\rangle \;=\; \partial_\mu V^\nu |e_\nu\rangle + V^\nu \partial_\mu |e_\nu\rangle \;=\; \Big(\partial_\mu V^\rho + V^\nu \Gamma^\rho_{\mu\nu}\Big)\, |e_\rho\rangle. \tag{6.16}$$

En la última igualdad hemos definido la *conexión de Levi-Civita* (o los *símbolos de Christoffel*) $\Gamma^\rho_{\mu\nu}$ como

$$\partial_\mu |e_\nu\rangle = \Gamma^\rho_{\mu\nu} |e_\rho\rangle, \tag{6.17}$$

o sea, $\Gamma^\rho_{\mu\nu}$ es el componente ρ del vector $\partial_\mu |e_\nu\rangle$ en la base $\{|e_\rho\rangle\}$. El objeto

$$\nabla_\mu V^\rho = \partial_\mu V^\rho + \Gamma^\rho_{\mu\nu} V^\nu \tag{6.18}$$

es la *derivada covariante* de V_μ y veremos que tiene la propiedad de que transforma como un tensor de rango 2 bajo cambio de coordenadas. Aunque no es obvio de la definición (6.17), la conexión de Levi-Civita $\Gamma^\rho_{\mu\nu}$ es simétrica en los índices μ y ν (ejerc.) y se puede expresar en función de la métrica como (ejerc.)

$$\Gamma^\rho_{\mu\nu} = \frac{1}{2} g^{\rho\lambda} \Big(\partial_\mu g_{\lambda\nu} + \partial_\nu g_{\mu\lambda} - \partial_\lambda g_{\mu\nu}\Big). \tag{6.19}$$

Es fácil de comprobar que la conexión de Levi-Civita es cero en coordenadas cartesianas, pero no lo es en coordenadas esféricas o cilíndricas. Utilizando estas definiciones de derivadas en coordenadas curvilíneas se puede obtener las expresiones conocidas del gradiente, la divergencia, el rotacional y el laplaciano en coordenadas esféricas y cilíndricas. Para su expresión referimos la literatura y para una discusión más elaborada sobre estos operadores en coordenadas arbitrarias al Capítulo 7.

6.3. Variedades

Hemos visto en la sección 5.4 que la física de la relatividad general sugiere que el espacio(tiempo) parece plano (Minkowski) a escalas pequeñas, pero globalmente no se puede describir con un sistema de coordenadas cartesianas. Matemáticamente el objeto que tiene esta propiedad es una *variedad* (en inglés *manifold*).

En grandes líneas se puede decir que una variedad N-dimensional $\mathcal{M}^N$ es un espacio que localmente tiene el aspecto de $\mathbb{R}^N$. En cada punto p de la variedad se puede definir el *espacio tangente en el punto p*, $T_p(\mathcal{M}^N)$, y ese espacio tangente es isomorfo a $\mathbb{R}^N$. Ahora, si decimos que la variedad es localmente $\mathbb{R}^N$, queremos decir que en una región pequeña alrededor de p, el espacio tangente $T_p(\mathcal{M}^N)$ es una buena aproximación a la variedad. Sin embargo, los espacios

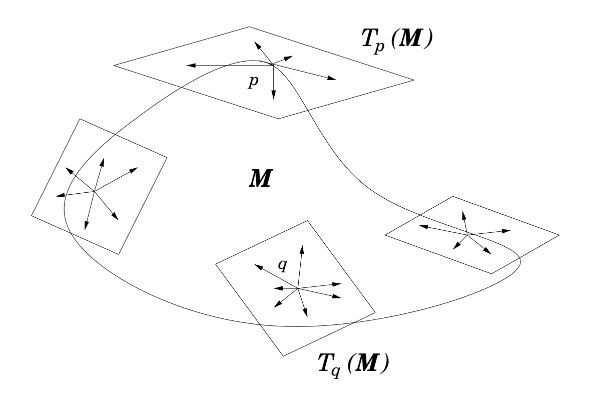


Figura 6.2: *El espacio tangente: el conjunto de todos los vectores en el punto p forma el espacio tangente $T_p(\mathcal{M})$ en p. El conjunto de los espacios tangentes en todos los puntos de la variedad se llama el haz tangente $T(\mathcal{M})$.*

tangentes en distintos puntos de la variedad son distintos, debido al hecho de que la variedad no es globalmente plana. El conjunto de todos los espacios tangentes de todos los puntos p de la variedad se llama el *haz tangente* $T(\mathcal{M}^N)$.

Un ejemplo concreto es el de un barco que navega entre dos puntos de la Tierra, sin darse cuenta que la Tierra es redonda. En cada momento el capitán sólo ve una región pequeña alrededor del barco y está tentado a pensar que la Tierra es plana, ya que identifica la región que ve con el plano tangente en el punto donde se encuentra. Según el barco va moviendo a lo largo de su ruta, va "saltando de plano tangente en plano tangente" y no se da cuenta de la curvatura de la Tierra mientras sólo hace medidas locales.

Una definición matemáticamente un poco más rigurosa del concepto de variedad sería decir que en cada punto de la variedad existe una transformación ϕ que mapea una parte $U \subset \mathcal{M}^N$ a $\mathbb{R}^N$ (véase Figura 6.3). Las coordenadas (no necesariamente cartesianas) en $\mathbb{R}^N$ inducen coordenadas locales x^μ en $\mathcal{M}^N$ a través de la transformación inversa ϕ^{-1}. Una parte U de $\mathcal{M}^N$ con un sistema de coordenadas inducidas se llama un *mapa*. De manera práctica se puede por lo tanto pensar en un mapa como una parte del espacio equipado con un sistema de coordenadas.

En general no será posible cubrir una variedad entera con sólo un mapa. $\mathbb{R}^N$ es un ejemplo trivial de una variedad que se puede cubrir con un solo mapa (en este caso la transformación ϕ es la unidad). Pero es imposible cubrir por ejemplo la esfera bidimensional $\mathbb{S}^2$ con un solo mapa. Este es el famoso Problema de los Cartógrafos: no existe una proyección tal que la Tierra entera aparece en un solo mapa plano. Por ejemplo, una proyección estereográfica desde el polo norte proyecta todos los puntos de $\mathbb{S}^2$ en el plano $\mathbb{R}^2$, salvo el mismo polo norte. Este sí se puede incluir si se hace la proyección estereográfica desde el polo sur, pero entonces no aparece este último. Para cubrir la 2-esfera completa hace falta un mímino de dos mapas.

Si dos mapas U_a y U_b se solapan en una región, las transformaciones inversas ϕ_a^{-1} y ϕ_b^{-1} inducen dos sistemas de coordenadas x^μ y y^α en los respectivos parches. En la región donde los parches se solapan, las transformaciones ϕ_a y ϕ_b (y por lo tanto las coordenadas inducidas) tienen que ser tales que exista en la región de solapamiento una transformación $\Psi = \phi_a \circ \phi_b^{-1}$ y su inversa que relacionen las coordenadas x^μ y y^α de cada mapa (véase figura 6.3). En otras palabras, en la región de solapamiento tiene que existir un cambio de coordenadas que nos permita relacionar expresiones escritas en un sistema de coordenadas con expresiones en el otro sistema. De esto modo las distintas "versiones locales de $\mathbb{R}^N$" de la variedad están "pegadas" de manera continua. Un conjunto de mapas que tiene esta propiedad y además cubre todo la variedad se llama un *atlas*.

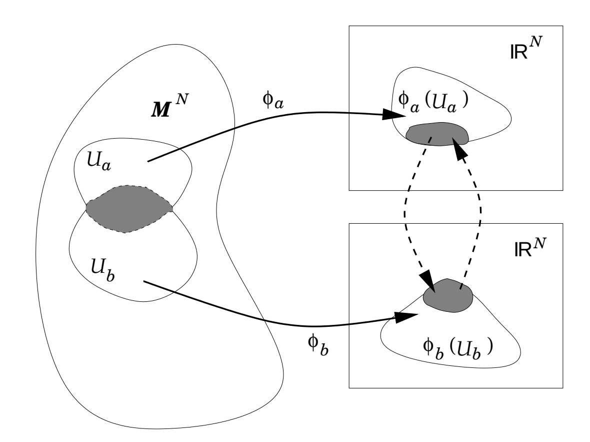


Figura 6.3: *Mapas y cambios de coordenadas en una variedad: un mapa es un parche U de una variedad N-dimensional $\mathcal{M}^N$ equipada con una transformación $\phi : U \to \mathbb{R}^N$ que induce un sistema de coordenadas en el mapa (flechas sólidas). Si dos mapas se solapan, los sistemas de coordenadas inducidas son tales que exista un cambio de coordenadas de un sistema al otro (las flechas interrumpidas).*

Finalmente, es útil distinguir las distintas variedades por la signatura de la métrica: una variedad se llama *riemanniana* si la métrica es definida positiva y *lorentziana* si tiene una métrica con signatura $(+ - - \ldots -)$. En el Capítulo 5 hemos visto que la coordenada que tiene el signo opuesto a las demás es una dirección temporal, de modo que variedades lorentzianas son espaciotiempos, mientras que variedades riemannianas sólo tienen direcciones espaciales y serán interpretadas como espacios (matemáticos).

Después de haber introducido el concepto de variedad en una manera más bien intuitiva,[6] vamos a estudiar las propiedades geométricas de estas variedades. En particular estaremos interesados en aquellas propiedades que no son válidas sólo en un sistema particular de coordenadas, sino en todas. Pero ya sabemos del Capítulo 4 que los objetos que tenemos que utilizar entonces para describir estas propiedades son los que transforman bien bajo cambios de coordenadas, es decir: los escalares, los tensores y los vectores.

6.4. Álgebra tensorial en variedades

En la sección 6.2 hemos introducido los cambios generales de coordenadas en $\mathbb{R}^N$ y obtenido una forma para la métrica, el producto escalar y incluso hemos introducido brevemente el operador diferencial y indicado las sutilezas que aparecen con este operador bajo cambios generales de coordenadas. Sin embargo, mientras sólo estabamos haciendo cambios de coordenadas en $\mathbb{R}^N$, la métrica $g_{\mu\nu}$ satisface unas ligaduras específicas: existen unas funciones $x^i(y^\mu)$ tal que la métrica se reduce a la métrica euclídea en coordenadas cartesianas δ_{ij} a través de la expresión (6.10).

En general en un variedad arbitraria $\mathcal{M}^N$ no existirá un cambio de coordenadas de esta forma. Por ejemplo, es imposible encontrar un cambio de coordenadas tal que la métrica de la 2-esfera $\mathbb{S}^2$,

$$ds^2 = R_0^2\Big(d\theta^2 + \sin^2\theta d\varphi^2\Big) \tag{6.20}$$

o del hiperboloide bidimensional $\mathbb{H}^2$,

$$ds^2 = R_0^2\Big(d\chi^2 + \sinh^2\chi d\xi^2\Big) \tag{6.21}$$

[6]Para una definición matemática más rigurosa referimos a R. Wald, *General Relativity* (1984) y M. Nakahara, *Geometry, Topology and Physics* (2003).

reduzcan a la métrica plana en todo la variedad. Parece por lo tanto que $\mathbb{S}^2$ y $\mathbb{H}^2$ tienen propiedades intrínsecas (i.e. que no dependen del sistema de coordenadas usado) distintas de $\mathbb{R}^N$. En el Capítulo 7 estudiaremos estas propiedades en detalle, pero primero tenemos que mirar cómo se comportan vectores y tensores bajo cambios generales de coordenadas en variedades arbitrarias, o sea en espacios que no son necesariamente $\mathbb{R}^N$. En realidad, en la sección anterior no hemos utilizado en ningún momento el hecho de que la métrica sea reducible o no a la cartesiana, de modo que todas las propiedades derivadas en la sección 6.2 (salvo precisamente la ecuación (6.10)) serán directamente generalizables a $\mathcal{M}^N$.

Supongamos por lo tanto que en una variedad N-dimensional $\mathcal{M}^N$ tenemos dos sistemas de coordenadas x^μ y y^α (donde utilizamos los índices griegos del principio del alfabeto para unos y los del medio del alfabeto para otros), que están relacionadas a través de un cambio general de coordenadas del tipo (6.3):

$$y^\alpha = y^\alpha(x^\mu) \qquad \Longleftrightarrow \qquad x^\mu = x^\mu(y^\alpha). \tag{6.22}$$

En contraste con el caso de una transformación global, las coordenadas x^μ e y^α no transforman como vectores bajo una transformación local. Efectivamente, el cambio de coordenadas (6.22) no es de la forma (4.5). Sin embargo, sus diferenciales

$$dy^\alpha = \frac{\partial y^\alpha}{\partial x^\mu}\, dx^\mu, \qquad dx^\mu = \frac{\partial x^\mu}{\partial y^\alpha}\, dy^\alpha \tag{6.23}$$

sí transforman como vectores, si identificamos las matrices de la transformación con las derivadas parciales

$$M^\alpha{}_\mu = \frac{\partial y^\alpha}{\partial x^\mu}. \tag{6.24}$$

Observa que las entradas de la matriz $M^\alpha{}_\mu$ no son constantes, sino funciones de x^μ, por el carácter local de la transformación. En general llamaremos un vector contravariante a cualquier objeto V^μ, que bajo un cambio de coordenadas (6.3) transforma como

$$V^\alpha = \frac{\partial y^\alpha}{\partial x^\mu} V^\mu. \tag{6.25}$$

Nótese que las transformaciones ortogonales (4.5) y (5.10) son casos especiales de este definición.

Matemáticamente un vector contravariante localizado en un punto p de la variedad no es un objeto que vive en la variedad, sino en el *espacio tangente* $T_p(\mathcal{M})$ del punto p. Estamos acostumbrados a pensar en vectores como objectos que conectan dos puntos del espacio. Esta idea es correcta para $\mathbb{R}^N$, pero no para una variedad arbitraria, por el hecho de que no es posible definir curvas preferidas entre dos puntos de una variedad sin suponer ciertas condiciones sobre las propiedades geométricas.[7]

En analogía con (4.16) y (6.25), un vector covariante es un objeto W_μ que transforma bajo (6.3) como

$$W_\alpha = \frac{\partial x^\mu}{\partial y^\alpha} W_\mu, \tag{6.26}$$

o sea, que transforma con la matriz inversa $(M^{-1})^\mu{}_\alpha = \partial x^\mu / \partial y^\alpha$. Matemáticamente un vector covariante es una uno-forma, un operador que actúa sobre vectores (contravariantes) para dar un escalar. Dicho de otra manera: el producto escalar de un vector covariante y uno contravariante es un escalar. Los vectores covariantes son elementos del *espacio cotangente* $T_p^*(\mathcal{M})$ o el *espacio dual*.

Sabiendo como transforman los vectores co- y contravariantes bajo cambios generales de coordenadas, es obvio generalizar las definiciones a tensores de rango (m, n). Un tensor de rango

[7] La razón por qué la idea sí es correcta en $\mathbb{R}^N$ es porque en cada punto su espacio tangente es el mismo $\mathbb{R}^N$, de modo que podemos identificar los distintos espacios tangentes y la variedad misma.

(m, n) está definido como un objeto que transforma bajo cambios generales de coordenadas como

$$T^{\mu_1 \ldots \mu_m}{}_{\nu_1 \ldots \nu_n} = \frac{\partial x^{\mu_1}}{\partial y^{\alpha_1}} \cdots \frac{\partial x^{\mu_m}}{\partial y^{\alpha_m}} \frac{\partial y^{\beta_1}}{\partial x^{\nu_1}} \cdots \frac{\partial y^{\beta_n}}{\partial x^{\nu_n}} T^{\alpha_1 \ldots \alpha_m}{}_{\beta_1 \ldots \beta_n} \tag{6.27}$$

Del mismo modo está claro que todas las operaciones algebráicas con vectores y tensores definidas en la sección 4.6 son fácilmente generalizables.

En la sección 6.2 hemos introducido la métrica a través del producto escalar de los vectores de base. También se puede obtener la expresión para $g_{\alpha\beta}$ a través del cambio de coordenadas (6.22). Si $g_{\mu\nu}$ es la expresión de la métrica en las coordenadas x^μ, el elemento de línea en las coordenadas y^α viene dado por

$$ds^2 \;=\; g_{\mu\nu} dx^\mu dx^\nu \;=\; g_{\mu\nu} \frac{\partial x^\mu}{\partial y^\alpha} \frac{\partial x^\nu}{\partial y^\beta} dy^\alpha dy^\beta \;\equiv\; g_{\alpha\beta} dy^\alpha dy^\beta, \tag{6.28}$$

donde en la segunda igualdad hemos utilizado las ecuaciones (6.23). Nótese que la relación entre $g_{\alpha\beta}$ y $g_{\mu\nu}$

$$g_{\alpha\beta} \;=\; \frac{\partial x^\mu}{\partial y^\alpha} \frac{\partial x^\nu}{\partial y^\beta} g_{\mu\nu} \tag{6.29}$$

nos dice que la métrica es un tensor de rango $(0, 2)$, como en el caso de las transformaciones globales. Además es casi igual que la expresión (6.10), sólo que aquí en general no será posible relacionar la métrica $g_{\alpha\beta}$ con la plana δ_{ij} ($\eta_{\mu\nu}$).

También aquí la métrica juega el mismo papel que la métrica plana: con ella y su inversa podemos contraer tensores, subir y bajar indices y convertir vectores covariantes en contravariantes y vice versa:

$$V_\mu = g_{\mu\nu} V^\nu, \qquad\qquad V^\nu = g^{\mu\nu} V_\nu. \tag{6.30}$$

El producto escalar entre dos vectores por lo tanto está definida de la manera usual y es independiente del sistema de coordenadas utilizado

$$V_\mu W^\mu \;=\; g_{\mu\nu} V^\nu W^\mu \;=\; \frac{\partial y^\alpha}{\partial x^\mu} \frac{\partial y^\beta}{\partial x^\nu} g_{\alpha\beta} \frac{\partial x^\mu}{\partial y^\gamma} V^\gamma \frac{\partial x^\nu}{\partial y^\delta} W^\delta \;=\; g_{\alpha\beta} V^\alpha W^\beta \;=\; V_\alpha W^\alpha, \tag{6.31}$$

donde en la penúltima igualdad hemos utilizado que por la regla de la cadena

$$\frac{\partial y^\alpha}{\partial x^\mu} \frac{\partial x^\mu}{\partial y^\gamma} = \delta^\alpha_\gamma. \tag{6.32}$$

6.5. Integración y elementos de volumen invariantes

De $\mathbb{R}^N$ sabemos que el elemento de volumen en coordenadas curvilíneas no es simplemente el producto de los dx^μ, sino que va multiplicado con la raíz cuadrada del determinante de la métrica, $\sqrt{|g|}$. Por ejemplo, en $\mathbb{R}^3$ en coordenadas cartesianas, esféricas y cilíndricas tenemos que

$$dx\, dx\, dz \;=\; r^2 \sin\theta\, dr\, d\theta\, d\varphi \;=\; \rho\, d\rho\, d\varphi\, dz. \tag{6.33}$$

En general, en una variedad arbitraria, la medida de una integral está definida de la misma manera,

$$\int d^n x \sqrt{|g|}\, f(x). \tag{6.34}$$

La razón por qué es preciso definir el elemento de volumen con el factor $\sqrt{|g|}$ es que el simple producto $d^n x = dx^1\, dx^2 \ldots dx^N$ no es invariante bajo cambios generales de coordenadas. Esto se ve si nos damos cuenta de que podemos escribir $d^n x$ a través del tensor de Levi-Civita como

$$d^n x \;=\; \frac{1}{N!} \varepsilon_{\mu_1 \ldots \mu_N}\, dx^{\mu_1} \ldots dx^{\mu_N}, \tag{6.35}$$

donde el factor $1/N!$ corrige la multiplicidad en la suma sobre los índices repetidos. Ahora, tomando en cuenta que el tensor de Levi-Civita es en realidad una densidad tensorial y transforma como (recuerda (4.61))

$$\varepsilon_{\mu_1 \ldots \mu_N} \;=\; \left|\frac{\partial x}{\partial y}\right| \frac{\partial y^{\alpha_1}}{\partial x^{\mu_1}} \cdots \frac{\partial y^{\alpha_N}}{\partial x^{\mu_N}}\, \varepsilon_{\alpha_1 \ldots \alpha_N} \tag{6.36}$$

donde $|\partial x/\partial y| = \det(\partial x^\mu/\partial y^\alpha)$ es el jacobiano, el determinante del cambio de coordenadas, no es difícil ver que $d^n x$ transforma como (ejerc.)

$$\begin{aligned} d^n x &= \frac{1}{N!}\, \varepsilon_{\mu_1 \ldots \mu_N}\, dx^{\mu_1} \ldots dx^{\mu_N} \\ &= \frac{1}{N!} \left|\frac{\partial x}{\partial y}\right| \varepsilon_{\alpha_1 \ldots \alpha_N}\, dy^{\alpha_1} \ldots dy^{\alpha_N} \\ &= \left|\frac{\partial x}{\partial y}\right| d^n y. \end{aligned} \tag{6.37}$$

En otras palabras, el simple producto de diferenciales $d^n x = dx^1\, dx^2 \ldots dx^N$ no transforma como un escalar, sino como una densidad de peso 1, debido a la presencia del tensor de Levi-Civita. Eso en principio es un problema, ya que no se puede usar para definir una medida para integrales, ya que no es invariante.

Sin embargo, sí podemos construir una medida invariante, multiplicándolo con otra densidad con el peso opuesto. Al tomar el determinante de la regla de transformación (6.29), vemos que el determinante de la métrica tampoco transforma como un escalar, sino como una densidad con peso (-2),

$$\det g(y) \;=\; \left|\frac{\partial x}{\partial y}\right|^2 \det g(x), \tag{6.38}$$

o equivalentemente,

$$\det g(x) \;=\; \left|\frac{\partial x}{\partial y}\right|^{-2} \det g(y). \tag{6.39}$$

Combinando las reglas de transformación (6.37) y (6.39), se ve fácilmente que la combinación invariante es

$$\sqrt{|g(x)|}\, d^n x \;=\; \sqrt{|g(y)|}\, d^n y, \tag{6.40}$$

y esa es efectivamente lo que aparece como medida en los integrales en variedades arbitrarias.

Aunque el factor $\sqrt{|g|}$ tiene poco significado físico, su presencia es clave a la hora de conseguir resultados físicamente acpetables. Veremos más adelante que al variar una acción

$$S \;=\; \int d^n x \sqrt{|g|}\, \mathcal{L}(g_{\mu\nu}, \phi) \tag{6.41}$$

con $\mathcal{L}(g_{\mu\nu}, \phi)$ la densidad lagrangiana, es precisamente la presencia del factor $\sqrt{|g|}$ que hace que los resultados sean covariantes.

6.6. Ejemplo concreto: cambios de coordenadas en $\mathbb{S}^2$

Considera la esfera bidimensional $\mathbb{S}^2$ con radio R_0, parametrizada de la manera usual por el ángulo polar θ y el ángulo azimutal φ. La métrica en estas coordenadas viene dada por la conocida expresión

$$ds^2 \;=\; R_0^2\Big(d\theta^2 \;+\; \sin^2\theta d\varphi^2\Big). \tag{6.42}$$

Obsérvese que este elemento de línea reproduce exactamente las expresiones conocidas para la distancia entre dos puntos en $\mathbb{S}^2$: la distancia entre dos puntos con la misma longitud ($\Delta\varphi = 0$)

es el segmento de arco $\Delta\theta$ multiplicado por el radio de la esfera, mientras la distancia entre dos puntos con la misma latitud ($\Delta\theta = 0$) viene dada por el segmento de arco $\Delta\varphi$, multiplicado por el radio $R_0 \sin\theta$ del círculo de latitud.

Consideramos ahora el siguiente cambio de coordenadas y su inversa

$$r = \frac{R_0 \sin\theta}{1-\cos\theta} \qquad \Longleftrightarrow \qquad \theta = 2 \operatorname{arc\,tg}\left(\frac{R_0}{r}\right). \tag{6.43}$$

Este cambio de coordenadas es un rescaleo del ángulo polar θ, que está bien definido en todos los puntos, salvo en el polo norte $\theta = 0$. Obsérvese que el polo sur, $\theta = \pi$, se corresponde con $r = 0$, mientras el ecuador, $\theta = \pi/2$ se corresponde con $r = R_0$. Sin embargo r tiende a infinito en el límite en que θ tiende a cero y el cambio de coordenadas mismo es divergente en este límite.

Para calcular la forma de la métrica en las coordenadas (r, φ), utilizaremos la expresión (6.28). De (6.43) tenemos directamente que

$$d\theta = \frac{-2R_0}{R_0^2 + r^2} dr, \qquad \sin\theta = \frac{2R_0 r}{R_0^2 + r^2}, \tag{6.44}$$

donde en la última expresión hemos utilizado la fórmula $\sin(2\alpha) = 2 \operatorname{tg}\alpha / (1 + \operatorname{tg}^2\alpha)$. Por lo tanto, la expresión para la métrica de la $\mathbb{S}^2$ en las nuevas coordenadas es (ejerc.)

$$ds^2 = \frac{4R_0^4}{(R_0^2 + r^2)^2} \Big[dr^2 + r^2 d\varphi^2\Big]. \tag{6.45}$$

Obsérvese que la parte entre paréntesis es la métrica plana de $\mathbb{R}^2$ en coordenadas polares, pero el factor no-trivial delante hace que la métrica total no sea la de $\mathbb{R}^2$. En el Capítulo 7 desarrollaremos técnicas para determinar la curvatura de variedades, pero a estas alturas ya debería quedar claro que la métrica (6.45) no puede ser la de $\mathbb{R}^2$, puesto que no hemos hecho nada más que aplicar un cambio de coordenadas a la esfera bidimensional, así que de algún modo la métrica (6.45) debe seguir describiendo la esfera (en el siguiente capítulo veremos cómo).

Sin embargo, sí hay una explicación de por qué aparece aquí algo tan parecido a la métrica de $\mathbb{R}^2$ en coordenadas polares. El cambio de coordenadas (6.43) tiene la interpretación geométrica de una proyección estereográfica desde el polo norte a un plano que pasa por el ecuador, y las coordenadas (r, φ) son las coordenadas polares en ese plano. La métrica (6.45), por lo tanto, no es la métrica euclídea usual en el plano, sino la métrica de la $\mathbb{S}^2$ inducida en el plano por la proyección estereográfica. Las propiedades geométricas del plano con la métrica (6.45) son muy diferentes a las del plano usual. Por ejemplo, dos puntos cercanos al polo sur se proyectan a puntos cerca del origen del plano $r = 0$, pero dos puntos cercanos al polo norte se proyectan a puntos lejos del origen y lejos uno del otro (con "lejos" queremos decir lejos en términos de la métrica usual en $\mathbb{R}^2$). Sin embargo, el prefactor de la métrica (6.45) disminuye muy rápido para r grande, de modo que la distancia medida por la métrica inducida es la misma entre ambos pares de puntos.

Aunque topológicamente las coordenadas (r, φ) describen un plano, geométricamente la métrica (6.45) todavía es la de $\mathbb{S}^2$.[8] En el siguiente capítulo estudiaremos que lo que sitingue los espacios con diferentes propiedades geométricas es la curvatura.

[8]El hecho de que el cambio de coordenadas (6.43) sea singular en el polo norte es una ilustración de que la topología de una esfera es la de un plano más un punto en el infinito.

Capítulo 7

Conexión afín y curvatura

En mi vida he trabajado tan duramente: he llegado a tener un gran respeto por las matemáticas, cuyas partes más sutiles había considerado hasta ahora ingenuamente como puro lujo. Comparado con este problema, la teoría original de la relatividad [especial] es un juego de niños.
(A. Einstein 1912, en una carta a A. Sommerfeld)

En los siguientes dos capítulos entraremos en el corazón de la matemática de tensores: la geometría diferencial o el cálculo de tensores. El problema central que de trata aquí es cómo construir operadores diferenciales en variedades arbitrarias de una manera covariante y cómo definir la curvatura de una variedad. En este capítulo trataremos el problema en toda generalidad, para conexiones arbitrarias, y en el Capítulo 8 miraremos específicamente a la conexión de Levi-Civita, relevante para la relativida general.

7.1. Transporte paralelo y la conexión afín

Ya hemos visto en la sección 6.2 que hay sutilezas que aparecen cuando dejamos actuar el operador diferencial $\partial_\mu = \partial/\partial x^\mu$ sobre vectores y tensores en un sistema de coordenadas curvilíneas en $\mathbb{R}^N$. Al variar de punto en punto la dirección de los vectores de base $|e_\mu\rangle$, aparecieron de manera natural los símbolos de Christoffel $\Gamma^\rho_{\mu\nu}$, como las derivadas de éstos. Ahora veremos que este problema no sólo se pone en $\mathbb{R}^N$, pero en general en cualquier variedad equipada con coordenadas curvilíneas y daremos una descripción independiente de una base y con un poco más de rigor matemático.

De manera análoga al caso de las transformaciones ortogonales, se puede demostrar a través de la regla de la cadena que la derivada de un campo escalar $\partial_\mu \phi \equiv \partial\phi/\partial x^\mu$ transforma como un vector covariante bajo cambios generales de coordenadas,

$$\partial_\alpha \phi = \frac{\partial \phi}{\partial y^\alpha} = \frac{\partial \phi}{\partial x^\mu}\frac{\partial x^\mu}{\partial y^\alpha} = \frac{\partial x^\mu}{\partial y^\alpha}\partial_\mu \phi. \tag{7.1}$$

Por otro lado, con la definición del operador ∂^μ hay que tener un poco más de cuidado. Dado que en las coordenadas x^μ no transforman como un vector bajo cambios generales de coordenadas, la cantidad x_μ no es el dual de x^μ y el operador $\partial/\partial x_\mu$ no está bien definido. Lo más sencillo es definir el operador dual a ∂_μ como $\partial^\mu \equiv g^{\mu\nu}\partial_\nu$. Está claro entonces que $\partial^\mu \phi$ se comporta como un vector contravariante.

Sin embargo, la derivada parcial de un vector o un tensor no transforma bien bajo cambios

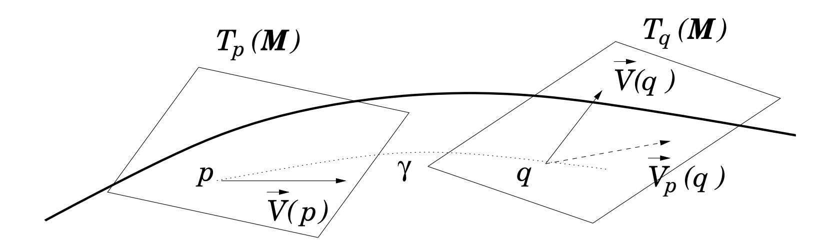


Figura 7.1: *Comparando vectores en distintos puntos de la variedad: Los vectores $|V(p)\rangle$ en p y $|V(q)\rangle$ en q viven en espacios tangentes diferentes, $T_p(\mathcal{M})$ y $T_q(\mathcal{M})$ respectivamente. Para poder compararlos, hay que definir un vector $|V_p(q)\rangle$ en q, que es el vector $|V(p)\rangle$ transportado paralelamente hasta q a lo largo de una curva γ.*

generales de coordenadas. Efectivamente, aplicando la regla de la cadena y usando que V^α transforma como un vector, tenemos que

$$\partial_\alpha V^\beta \;=\; \frac{\partial x^\mu}{\partial y^\alpha}\partial_\mu\Big(\frac{\partial y^\beta}{\partial x^\nu}V^\nu\Big) \;=\; \frac{\partial x^\mu}{\partial y^\alpha}\frac{\partial y^\beta}{\partial x^\nu}\,\partial_\mu V^\nu \;+\; \frac{\partial x^\mu}{\partial y^\alpha}V^\nu\frac{\partial^2 y^\beta}{\partial x^\mu \partial x^\nu}. \tag{7.2}$$

El primer término de la derecha sí tiene la forma adecuada para la transformación de un tensor de rango $(1,1)$, pero tenemos un término extra, que rompe el carácter tensorial de esta regla de transformación. Nótese que el término en cuestión es proporcional a la derivada de la matriz de la transformación $\partial_\mu M^\beta{}_\nu$. La presencia de este término es debido al carácter local de la transformación, ya que en el caso de una transformación global las entradas de la matriz $M^\beta{}_\nu$ son constantes.

La razón por qué en variedades arbitrarias la derivada parcial de un vector (o un tensor) no es un tensor, es debido al carácter local de las transformaciones. Más precisamente, si tomamos la derivada de un campo vectorial, estamos comparando el valor del campo en un punto p con coordenadas x^μ, con el valor del campo en un punto q, infinitesimalmente cerca de p, con coordenadas $x^\mu + \delta x^\mu$. Pero debido al carácter local del cambio de coordenadas, la matriz $M^\alpha{}_\nu = \partial y^\alpha/\partial x^\mu$ tiene un valor distinto en p que en q. No es de extrañar por lo tanto que

$$\lim_{\delta x^\nu \to 0} \frac{V^\mu(q) - V^\mu(p)}{\delta x^\nu} \tag{7.3}$$

no transforma bien, ya que las reglas de transformación (6.25) son distintas en los dos puntos. Efectivamente, notado que el término extra es proporcional a la derivada de la matriz de la transformación.

Matemáticamente hablando el problema es peor. Ya hemos dicho en la sección 6.4 que los vectores no viven en la variedad, sino en el espacio tangente. Si queremos comparar dos vectores en los puntos p y q, estamos en realidad comparando dos objetos que viven en espacios vectoriales distintos, $T_p(\mathcal{M})$ y $T_q(\mathcal{M})$ respectivamente (Véase Figura 7.1). Por lo tanto, matemáticamente la expresión (7.3) no tiene sentido.

Claramente, lo que necesitamos es un vector $V_p^\mu(q)$ que vive en $T_q(\mathcal{M})$ y que tiene toda la información sobre el vector $V^\mu(p)$ en $T_p(\mathcal{M})$. Dado que $V_p^\mu(q)$ y $V^\mu(q)$ viven en el mismo espacio vectorial, no habría ningún problema al comparar los dos vectores.

El vector $V_p^\mu(q)$ sería el *transportado paralelo de* $V^\mu(p)$. La idea es que en una variedad arbitraria, un campo vectorial puede cambiar de punto en punto por dos razones distintas: o bien por las variaciones dinámicas del campo campo (igual que en $\mathbb{R}^N$), o bien por los efectos de la variedad. Un ejemplo físico de lo primero sería un campo vectorial correspondiente a la dirección y fuerza del viento en cada punto de la Tierra. Las variaciones de este campo vectorial entonces son dinámicas, ya que son debidas a las distintas circunstancias meteorológicas en diferentes puntos.

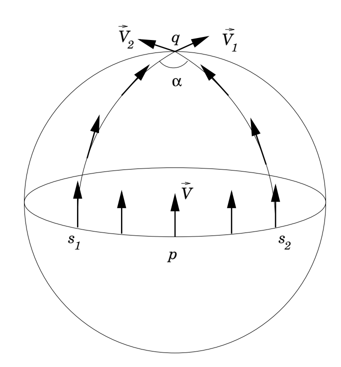


Figura 7.2: *En una variedad con curvatura,, el resultado del transporte paralelo de un vector de un punto p a un punto q depende de la curva entre p y q que uno usa para transportar el vector.*

Un ejemplo del segundo sería un campo vectorial constante apuntando al Polo Norte. Aquí también varía el campo vectorial de punto en punto, pero por razones geométricas (y no dinámicas): la dirección norte es distinta en diferentes puntos de la Tierra.

El transportado paralelo $V_p^\mu(q)$ sería por lo tanto el vector $V^\mu(p)$ en el punto q "si el campo vectorial no hubiera cambiado", es decir "si el cambio sólo hubiera sido causado por la variedad". Una vez hallado $V_p^\mu(q)$, el objeto

$$\lim_{\delta x^\nu \to 0} \frac{V^\mu(q) - V_p^\mu(q)}{\delta x^\nu} \tag{7.4}$$

estaría bien definido, puesto que $V^\mu(q)$ y $V_p^\mu(q)$ viven en el mismo espacio.

La pregunta inmediata que surge ahora es ¿cómo se define el vector $V_p^\mu(q)$? En otras palabras, ¿cómo se identifica el vector de $T_q(\mathcal{M})$ que es paralelo a $V^\mu(p)$ en $T_p(\mathcal{M})$? En realidad esta pregunta es imposible contestarla en esta forma, porque en una variedad arbitraria el resultado de un transporte paralelo de un vector de un punto p a otro q depende de la curva que uno sigue para ir de p a q (vease figura 7.2). Como veremos más adelante en la sección 7.3, esta propiedad es básicamente la definición de curvatura en una variedad. Para determinar cuál es el vector $V_p^\mu(q)$ en $T_q(\mathcal{M})$, tenemos que decir cómo hemos transportado $V^\mu(p)$ hasta el punto q.

Supongamos que transportamos el vector $V^\mu(p)$ hasta q a lo largo de una de las direcciones de coordenadas. Es decir, si el punto p tiene coordenadas x^μ y el punto q coordenadas $x^\mu + \delta x^\mu$ y trasladamos el vector $V^\mu(p)$ a lo largo del intervalo δx^μ, podemos escribir el nuevo vector $V_p^\mu(q)$ como $V_p^\mu(q) = V^\mu(p) + \delta V^\mu$. En otras palabras, el vector transportado paralelo es un vector original más un término (¡no tensorial!) de corrección, que depende de δx^μ. Es razonable suponer que para desplazamientos pequeños, el término de corrección es lineal tanto en δx^μ como en $V^\mu(p)$. Por lo tanto tenemos que

$$V_p^\mu(q) = V^\mu(p) - \Gamma^\mu_{\nu\rho}\delta x^\nu V^\rho(p), \tag{7.5}$$

donde los $\Gamma^\mu_{\nu\rho}$, que aparecen como los coefficientes de proporcionalidad, es la *conexión afín* o simplemente la *conexión*. El signo menos se ha elegido por futuro comodidad. La conexión es el conjunto de N^3 funciones (en una variedad N-dimensional) de las coordenadas x^μ que define cómo hacer el transporte paralelo entre los puntos. Concretamente, $\Gamma^\mu_{\nu\rho}$ es el cambio de la componente μ del vector en la dirección ρ bajo el desplazamiento en la dirección ν. Los $\Gamma^\mu_{\nu\rho}$'s que hemos encontrado en la sección 6.2 son un caso especial de la conexión introducido en (7.5), llamado la conexión de Levi-Civita, que definiremos en el siguiente capítulo.

Podemos ilustrar esto con el ejemplo de transporte paralelo en $\mathbb{R}^2$, con la noción intuitiva de paralelo, la conexión de Levi-Civita: las componentes cartesianas V^i no cambian bajo una traslación en una dirección x^j, así que en coordenadas cartesianas todos las componentes de la

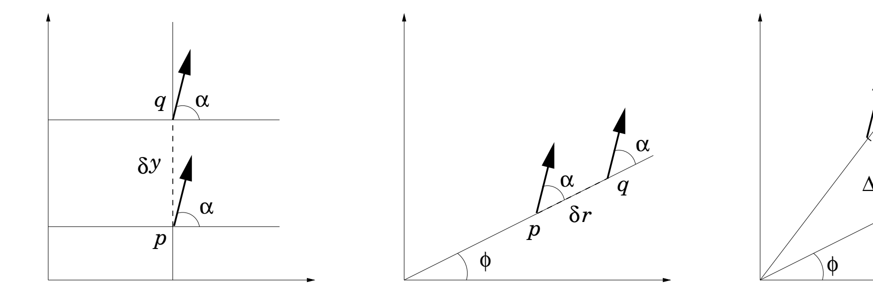


Figura 7.3: *El transporte paralelo en $\mathbb{R}^2$ (utilizando la conexión de Levi-Civita) en coordenadas cartesianas y polares. Las componentes del vector siguen inalteradas bajo un traslado en una dirección cartesiana. Sin embargo, al trasladar un vector $V^\mu(p)$ de p a q en una dirección radial cambia la componente V^φ. Si se traslada en la dirección φ, cambian tanto la componente V^r como V^φ.*

conexión son cero. Sin embargo, en coordenadas polares (r,φ) las componentes no nulas de la conexión son

$$\Gamma^\varphi_{r\varphi} = \Gamma^\varphi_{\varphi r} = r^{-1}, \qquad\qquad \Gamma^r_{\varphi\varphi} = -r, \tag{7.6}$$

lo que implica que bajo un desplazamiento en r el vector transportado paralelo tiene componentes

$$V^r_p(q) = V^r(p), \qquad\qquad V^\varphi_p(q) = V^\varphi(p) - \frac{1}{r}V^\varphi(p)\delta r, \tag{7.7}$$

y bajo un desplazamiento en la dirección φ

$$V^r_p(q) = V^r(p) + rV^\varphi(p)\delta\varphi, \qquad\qquad V^\varphi_p(q) = V^\varphi(p) - \frac{1}{r}V^r(p)\delta\varphi. \tag{7.8}$$

Dejamos como ejercicio comprobar que este resultado es efectivamente lo que sale al trasladar un vector radialmente de un punto p con coordenadas (r,φ) a uno punto q con coordenadas $(r+\delta r,\varphi)$, y angularmente a un punto q con coordenadas $(r,\varphi+\delta\varphi)$ respectivamente (véase Figura 7.3).

La pregunta de antes, sobre cómo identificar el vector de $T_q(\mathcal{M})$ que es paralelo a $V^\mu(p)$, se puede responder ahora de la siguiente manera: Dada una conexión $\Gamma^\rho_{\mu\nu}$ y dada una curva γ entre los puntos p y q, se identifica $V^\mu_p(q)$ a través de la formula (7.5). Pero la pregunta básica queda: ¿cuál es la conexión adecuada a utilizar en una variedad determinada? ¿Cómo conseguimos la forma de la N^3 funciones $\Gamma^\rho_{\mu\nu}$ para poder aplicar la formula (7.5)?

La respuesta otra vez es difícil. Resulta que en una variedad determinada hay muchas (¿infinitas?) conexiones posibles, que todas definen de manera distinta el vector $V^\mu_p(q)$. La eleción de qué conjunto de funciones $\Gamma^\rho_{\mu\nu}$ utilizamos es arbitraria. Esto es el resultado del hecho de que en realidad no hay una manera natural de comparar vectores de diferentes espacios vectoriales y que en variedades arbitrarias no existe una noción natural de paralelo.

Afortunadamente, si existe una métrica sobre nuestra variedad, podemos definir una conexión preferida, con ciertas propiedades que la hace única. Esta conexión es la *conexión de Levi-Civita* y la estudiaremos en más detalle en el Capítulo 8. Pero merece la pena enfatisar que las condiciones que satsiface la conexión de Levi-Civita para tener las mencionadas propiedades, son sobre todo condiciones físicas, no tanto matemáticas. Matemáticamente no hay ninguna razón para preferir la conexión de Levi-Civita ante cualquier otra conexión posible, salvo su unicidad. De momento seguiremos trabajando con una conexión general y en el siguiente capítulo nos centraremos en la conexión de Levi-Civita.

7.2. La derivada covariante

Supongamos ahora que tenemos una conexión $\Gamma^\rho_{\mu\nu}$, que nos dice cómo hacer transporte paralelo a vectores de un punto p a un punto q. Con esta estructura matemática, podemos definir un nuevo tipo de operador diferencial que no tiene el problema de la derivada parcial, es decir que transforma como un tensor bajo cambios generales de coordenadas.

Aplicando la fórmula (7.4), donde utilizamos la expresión (7.5) para $V_p^\mu(q)$ obtenemos un operdaor diferencial, llamada la *derivada covariante*:

$$\nabla_\nu V^\mu = \partial_\nu V^\mu + \Gamma^\mu_{\nu\rho} V^\rho. \tag{7.9}$$

La derivada covariante actúa por lo tanto como una parcial, pero luego suma un término de corrección, lineal en V^μ, debido al transporte paralelo del vector. Podemos generalizar el concepto de derivada covariante a lo largo de una direccion x^μ a una derivada covariante a lo largo de una curva $x^\mu(\tau)$ definiendo como $\nabla_\gamma V^\mu = u^\rho \nabla_\rho V^\mu$, donde $u^\mu = dx^\mu(\tau)/d\tau$ es el vector tangente a la curva $x^\mu(\tau)$. Decimos entonces que un vector V^μ es transportado paralelo a lo largo de una curva $x^\mu(\tau)$ si $u^\mu \nabla_\mu V^\nu = 0$ a lo largo de toda la curva.

Por otro lado, también es necesario definir derivadas covariantes para tensores, ya que con un razonamiento análogo al caso de vectores, la deriavda pacial de tensores tampoco transforma bien bajo cambios generales de coordenadas. Definimos por lo tanto la derivada covariante actuando sobre un tensor de rango $(2,0)$ como

$$\nabla_\mu T^{\nu\rho} \ = \ \partial_\mu T^{\nu\rho} \ + \ \Gamma^\nu_{\mu\lambda} T^{\lambda\rho} \ + \ \Gamma^\rho_{\mu\lambda} T^{\nu\lambda}. \tag{7.10}$$

Fíjese que la derivada covariante de un tensor de rango $(2,0)$ tiene dos términos de corrección, uno por cada índice, y que para cada índice dentro de su término de corrección, la estructura de los índices es igual que para una vector.

La derivada covariante (el operador ∇ se pronuncia *nabla*) comparte con la derivada parcial las propiedades de linealidad y el regla del producto de Leibniz

$$\begin{aligned} \nabla_\mu(\alpha V^\nu + \beta W^\nu) \ &= \ \alpha \nabla_\mu V^\nu + \beta \nabla_\mu W^\nu, \\ \nabla_\mu(V^\nu W^\rho) \ &= \ (\nabla_\mu V^\nu) W^\rho + V^\nu (\nabla_\mu W^\rho), \end{aligned} \tag{7.11}$$

pero encima tiene la propiedad que transforma bien como un tensor de rango $(1,1)$, por lo menos si elegimos bien las reglas de transformación de la conexión. Efectivamente, exigiendo que

$$\nabla_\alpha V^\gamma = \frac{\partial x^\mu}{\partial y^\alpha} \frac{\partial y^\gamma}{\partial x^\rho} \nabla_\mu V^\rho \tag{7.12}$$

y sustituyendo la definición (7.9) encontramos que

$$\begin{aligned} \partial_\alpha V^\gamma + \Gamma^\gamma_{\alpha\beta} V^\beta \ &= \ \frac{\partial x^\mu}{\partial y^\alpha} \frac{\partial y^\gamma}{\partial x^\rho} \Big(\partial_\mu V^\rho + \Gamma^\rho_{\mu\nu} V^\nu \Big) \\ &= \ \partial_\alpha V^\gamma \ - \ \frac{\partial x^\mu}{\partial y^\alpha} V^\rho \frac{\partial^2 y^\gamma}{\partial x^\mu \partial x^\rho} \ + \ \frac{\partial x^\mu}{\partial y^\alpha} \frac{\partial y^\gamma}{\partial x^\rho} \Gamma^\rho_{\mu\nu} V^\nu, \end{aligned} \tag{7.13}$$

donde en la segunda igualdad hemos utilizado las reglas (7.2) de transformación de $\partial_\mu V^\rho$. Quitando a la izquierda y a la derecha el término $\partial_\alpha V^\gamma$ y escribiendo el vector V^β en las coordenadas x^μ, obtenemos

$$\Gamma^\gamma_{\alpha\beta} \frac{\partial y^\beta}{\partial x^\nu} V^\nu \ = \ \frac{\partial x^\mu}{\partial y^\alpha} \frac{\partial y^\gamma}{\partial x^\rho} \Gamma^\rho_{\mu\nu} V^\nu - \frac{\partial x^\mu}{\partial y^\alpha} \frac{\partial^2 y^\gamma}{\partial x^\mu \partial x^\nu} V^\nu. \tag{7.14}$$

Dado que este relación tiene que ser verdad para cualquier vector V^ν, encontramos que las reglas de transformación de la conexión vienen dadas por

$$\Gamma^\gamma_{\alpha\beta} \ = \ \frac{\partial x^\mu}{\partial y^\alpha} \frac{\partial x^\nu}{\partial y^\beta} \frac{\partial y^\gamma}{\partial x^\rho} \Gamma^\rho_{\mu\nu} - \frac{\partial^2 y^\gamma}{\partial x^\mu \partial x^\nu} \frac{\partial x^\mu}{\partial y^\alpha} \frac{\partial x^\nu}{\partial y^\beta}. \tag{7.15}$$

Vemos por lo tanto que $\Gamma^\rho_{\mu\nu}$ no se comporta como un tensor de rango $(1,2)$, debido al segundo termino de la derecha. En cierto modo esto era de esperar: si exigimos que $\nabla_\mu V^\rho$ transforme como un tenor y sabemos que $\partial_\mu V^\rho$ no lo hace, es preciso que los $\Gamma^\rho_{\mu\nu}$ transformen con un término no-tensorial que cancela el término no-tensorial en la transformación de $\partial_\mu V^\rho$. Obsérvese sin embargo, que el término no-tensioral de las reglas de transformación (7.15) es independiente de $\Gamma^\rho_{\mu\nu}$, de modo que la diferencia $K^\rho_{\mu\nu} = \Gamma^\rho_{\mu\nu} - \tilde{\Gamma}^\rho_{\mu\nu}$ entre dos conexiones sí transforma como un tensor.

En la sección anterior hemos visto en $\mathbb{R}^2$ que en coordenadas cartesianas los Γ^k_{ij} son todos cero, mientras en coordenadas polares los $\Gamma^\rho_{\mu\nu}$ en general no lo son. Esto es una consecuencia del carácter no-tensorial de la conexión. Si los $\Gamma^\rho_{\mu\nu}$ transformase como un tensor y todos los componentes fueran cero en coordenadas cartesianas, serían también cero en todos los sistemas de coordenadas, lo que obviamente no es el caso.

En (7.9) hemos construido la derivada covariante de un vector contravariante V^μ. Para ver cómo actúa la derivada covariante en un vector covariante W_μ, es suficiente darnos cuenta de que la derivada covariante de un escalar ϕ es simplemente la derivada parcial:

$$\nabla_\mu \phi = \partial_\mu \phi. \tag{7.16}$$

Podemos por lo tanto construir un escalar ϕ de la contracción de un vector covariante y uno contravariante $\phi = W_\mu V^\mu$ y exigir que la derivada covariante actuando sobre ϕ es lo mismo que la derivada parcial. Supongamos que la derivada covariante de W_μ es de la forma $\nabla_\mu W_\nu = \partial_\mu W_\nu + \tilde{\Gamma}^\rho_{\mu\nu} W_\rho$, donde los N^3 funciones $\tilde{\Gamma}^\rho_{\mu\nu}$ quedan por determinar. Entonces tenemos que

$$\begin{aligned}
\nabla_\mu (W_\nu V^\nu) &= (\nabla_\mu W_\nu) V^\nu + W_\nu (\nabla_\mu V^\nu) \\
&= (\partial_\mu W_\nu + \tilde{\Gamma}^\rho_{\mu\nu} W_\rho) V^\nu + W_\nu (\partial_\mu V^\nu + \Gamma^\nu_{\mu\rho} V^\rho) \\
&= \partial_\mu (W_\nu V^\nu) + \tilde{\Gamma}^\rho_{\mu\nu} W_\rho V^\nu + \Gamma^\rho_{\mu\nu} W_\rho V^\nu \\
&\equiv \partial_\mu (W_\nu V^\nu).
\end{aligned} \tag{7.17}$$

Esto claramente sólo puede ser verdad si $\tilde{\Gamma}^\rho_{\mu\nu} = -\Gamma^\rho_{\mu\nu}$, o en otras palabras, si la derivada covariante actuando sobre un vector covariante es de la forma

$$\nabla_\mu W_\nu = \partial_\mu W_\nu - \Gamma^\rho_{\mu\nu} W_\rho. \tag{7.18}$$

Sabiendo cómo actúa una derivada covariante sobre vectores co- y contravariantes, podemos definir cómo actúa sobre de rango $(1,1)$ si construimos un tensor $T_\mu{}^\nu = V_\mu W^\nu$. No es difícil ver que la derivada covariante viene dada por

$$\nabla_\rho T_\mu{}^\nu = \partial_\rho T_\mu{}^\nu - \Gamma^\lambda_{\rho\mu} T_\lambda{}^\nu + \Gamma^\nu_{\rho\lambda} T_\mu{}^\lambda. \tag{7.19}$$

En otras palabras, la derivada covariante actuando sobre un tensor de rango $(1,1)$ es la derivada parcial más un término de corrección por cada índice, con los signos apropiados según el caso. En general entonces la derivada covariante de un tensor de rango (m,n) viene dada por

$$\begin{aligned}
\nabla_\rho T^{\mu_1 \ldots \mu_m}{}_{\nu_1 \ldots \nu_n} &= \partial_\rho T^{\mu_1 \ldots \mu_m}{}_{\nu_1 \ldots \nu_n} + \Gamma^{\mu_1}_{\rho\lambda} T^{\lambda \mu_2 \ldots \mu_m}{}_{\nu_1 \ldots \nu_n} + \ldots \\
&\quad + \Gamma^{\mu_m}_{\rho\lambda} T^{\mu_1 \ldots \mu_{m-1}\lambda}{}_{\nu_1 \ldots \nu_n} - \Gamma^\lambda_{\rho\nu_1} T^{\mu_1 \ldots \mu_m}{}_{\lambda \nu_2 \ldots \nu_n} - \ \ldots \ - \Gamma^\lambda_{\rho\nu_n} T^{\mu_1 \ldots \mu_m}{}_{\nu_1 \ldots \nu_{n-1}\lambda}.
\end{aligned} \tag{7.20}$$

Por último, dada una curva $x^\mu(\tau)$ en la variedad, entonces un tensor $T^{\mu_1 \cdots \mu_m}{}_{\nu_1 \ldots \nu_n}$ es transportado paralelo a lo largo de la curva, si la derivada covariante del tensor a lo largo de la curva es cero,

$$u^\rho \nabla_\rho T^{\mu_1 \cdots \mu_m}{}_{\nu_1 \ldots \nu_n} = 0, \tag{7.21}$$

donde $u^\mu = dx^\mu / d\tau$ es el vector tangente a la curva $x^\mu(\tau)$.

7.3. Curvatura y el tensor de Riemann

Intuitivamente tenemos una idea más o menos vaga de lo que es curvatura. Nos podemos imaginar una superficie bidimensional curva, pero un espacio tridimensional con curvatura ya es mucho más difícil. Intuitivamente estamos acostumbrados de pensar en (hiper)superficies curvos como embebidas en un espacio plano de dimensión más grande, pero en esta sección veremos que el concepto de variedades con curvatura existe independientemente de una inmersión en un espacio más grande. En otras palabras, una hormiga bidimensional puede distinguir si la superficie que recorre tiene curvatura o no, estudiando la geometría de la variedad, sin tener ni idea si esta superficie está embebida en $\mathbb{R}^3$.

Ya hemos dicho antes que en general el resultado del transporte paralelo de un vector de un punto p a un punto q depende de la trayectoria entre p y q. Por ejemplo, en una 2-esfera $\mathbb{S}^2$, podemos trasladar un vector $\vec{V}$ de un punto p en el ecuador al punto q en el polo norte, o bien primero un tramo al oeste hasta un punto s_1 y luego hacia el norte, o bien primero hacia el este hasta un punto s_2 y luego hacia el norte (véase Figura 7.2). Los vectores resultantes, $\vec{V}_1$ y $\vec{V}_2$, difieren con un ángulo α, igual al ángulo entre las dos trayectorias dirigidas hacia el polo norte.

El hecho de que los vectores no coincidan es una manifestación de la curvatura de la variedad encerrada por las curvas. Básicamente es la definición de curvatura: *diremos que una variedad es curva si el transporte paralelo de un vector a través de una curva cerrada resulta en un vector diferente al volver en el punto de salida.* Con esta definición está claro que la curvatura de una variedad tiene todo que ver con la manera cómo hacemos el transporte paralelo en la variedad y por lo tanto con la conexión $\Gamma_{\mu\nu}^{\rho}$ que hemos elegido.

Para definir bien, es decir de manera cuántitativa, la curvatura de una variedad consideramos el transporte paralelo de un vector a lo largo de un paralelogramo infinitesimal cuyos lados consisten de los vectores dx^μ y dx^ν, un desplazamiento infinitesimal en las direcciones x^μ y x^ν respectivamente. En particular, calculamos la diferencia entre trasladar el vector V^λ primero a lo largo de dx^μ y luego a lo largo de dx^ν y transladarlo primero dx^ν y después dx^μ. En otras palabras, tenemos que calcular el conmutador de las derivadas covariantes $[\nabla_\mu, \nabla_\nu] = \nabla_\mu\nabla_\nu - \nabla_\nu\nabla_\mu$ actuando sobre el vector V^λ.

En un primer paso calculamos

$$\begin{aligned}\nabla_\mu\nabla_\nu V^\lambda &= \partial_\mu\partial_\nu V^\lambda + \partial_\mu\Gamma_{\nu\rho}^{\lambda}V^\rho + \Gamma_{\nu\rho}^{\lambda}\partial_\mu V^\rho \\ &\quad -\Gamma_{\mu\nu}^{\rho}\partial_\rho V^\lambda - \Gamma_{\mu\nu}^{\rho}\Gamma_{\rho\sigma}^{\lambda}V^\sigma + \Gamma_{\mu\rho}^{\lambda}\partial_\nu V^\rho + \Gamma_{\mu\rho}^{\lambda}\Gamma_{\nu\sigma}^{\rho}V^\sigma,\end{aligned} \tag{7.22}$$

de modo que el conmutador viene dado por (ejerc.)

$$[\nabla_\mu, \nabla_\nu]V^\lambda = R_{\mu\nu\rho}{}^{\lambda}V^\rho \; - \; T_{\mu\nu}^{\rho}\nabla_\rho V^\lambda, \tag{7.23}$$

donde

$$\begin{aligned}R_{\mu\nu\rho}{}^{\lambda} &= \partial_\mu\Gamma_{\nu\rho}^{\lambda} - \partial_\nu\Gamma_{\mu\rho}^{\lambda} + \Gamma_{\mu\sigma}^{\lambda}\Gamma_{\nu\rho}^{\sigma} - \Gamma_{\nu\sigma}^{\lambda}\Gamma_{\mu\rho}^{\sigma}, \\ T_{\mu\nu}^{\rho} &= \Gamma_{\mu\nu}^{\rho} - \Gamma_{\nu\mu}^{\rho}.\end{aligned} \tag{7.24}$$

Vemos que, en contraste con las derivadas parciales, el conmutador de dos derivadas covariantes no es cero, sino que consiste de dos partes. Una parte es proporcional a vector V^ρ, donde el factor de proporcionalidad $R_{\mu\nu\rho}{}^{\lambda}$ es el *tensor de Riemann*. Esta parte mide la diferencia entre el transporte paralelo de V^ρ por una trayectoria y por otra y es por lo tanto una medida de la curvatura del espacio encerrado en el paralelogramo (véase Figura 7.4).

La segunda parte es proporcional a la derivada covariante de V^λ y el factor de proporcionalidad $T_{\mu\nu}{}^{\rho}$ es la parte antisimétrica de la conexión, llamado el *tensor de torsión*. El tensor de torsión

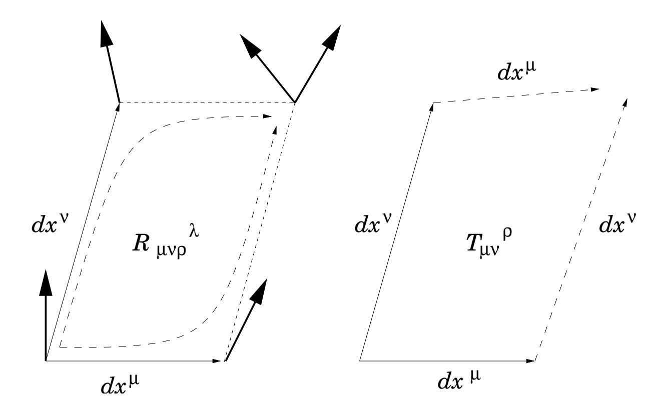


Figura 7.4: *El tensor de Riemann y el tensor de torsión: El tensor de Riemann $R_{\mu\nu\rho}{}^{\lambda}$ mide la diferencia en el transporte paralelo de un vector en un paralelograma infinitesimal y es una medida de la curvatura encerrada en el paralelograma. El tensor de torsión $T^{\rho}_{\mu\nu}$ es la parte antisimétrica de la conexión y mide hasta qué punto el paralelograma está cerrada.*

mide el grado en que cierre el paralelograma. Es decir, si transportamos paralelamente el vector infinitesimal dx^μ a lo largo de dx^ν, resultando en un vector dx'^μ y lo comparamos con el vector dx'^ν, siendo el transportado paralelo de dx^ν a lo largo de dx^μ, resulta que en general los vectores dx'^μ y dx'^ν no terminan en el mismo punto (vease figura 7.4). En otras palabras, en general el paralelograma dx^μ-dx^ν-dx'^μ-dx'^ν no necesariamente cierra. Sólo si nos restringimos a conexiones simétricas (una de las condiciones para la conexión de Levi-Civita), el tensor de torsión es cero y las trayectorias dx^μ-dx^ν y dx^ν-dx^μ formarán un paralelograma cerrado.

Con un mismo cálculo que en el caso del vector contravariante, podemos calcular el conmutador sobre escalares y tensores en general. Para un escalar ϕ tenemos que (ejerc.)

$$[\nabla_\mu, \nabla_\nu]\phi = -T^{\rho}_{\mu\nu}\nabla_\rho\phi, \tag{7.25}$$

y para un tensor $S^{\rho_1\cdots\rho_m}{}_{\lambda_1\ldots\lambda_n}$ (ejerc.)

$$\begin{aligned}[\nabla_\mu, \nabla_\nu]S^{\rho_1\cdots\rho_m}{}_{\lambda_1\ldots\lambda_n} &= R_{\mu\nu\sigma}{}^{\rho_1}S^{\sigma\rho_2\cdots\rho_m}{}_{\lambda_1\ldots\lambda_n} + \ \ldots\ + R_{\mu\nu\sigma}{}^{\rho_m}S^{\rho_1\cdots\sigma}{}_{\lambda_1\ldots\lambda_n} \\ &-R_{\mu\nu\lambda_1}{}^{\sigma}S^{\rho_1\cdots\rho_m}{}_{\sigma\lambda_2\ldots\lambda_n} - \ \ldots\ - R_{\mu\nu\lambda_n}{}^{\sigma}S^{\rho_1\cdots\rho_m}{}_{\lambda_1\ldots\sigma} - T^{\sigma}_{\mu\nu}\nabla_\sigma S^{\rho_1\cdots\rho_m}{}_{\lambda_1\ldots\lambda_n}.\end{aligned} \tag{7.26}$$

Como ya lo indican sus nombres, el tensor de Riemann y el tensor de torsión son tensores de rango $(1,3)$ y $(1,2)$ respectivamente, a pesar de que estén construidos a base de objetos que no son tensoriales (derivadas parciales y conexiones). Se puede comprobar el carácter tensorial de $R_{\mu\nu\rho}{}^{\lambda}$ y $T^{\rho}_{\mu\nu}$ calculando sus reglas de transformación a lo bruto, pero es más fácil darse cuenta que todos los elementos del lado izquierdo de (7.23) son tensores, así que el lado derecho también tiene que transformar como un tensor: dado que la torsión es un tensor por ser la diferencia entre dos conexiones, también el tensor de Riemann tiene que tener un carácter tensorial, para el lado derecha de (7.23) transforme bien.

Como ya hemos dicho, el tensor de Riemann mide la variación de un vector tras transporte paralelo alrededor de una curva cerrada. En particular, si el tensor de Riemann es cero, el transporte paralelo de un vector no depende de la trayectoria, lo que implica que la variedad es plana. También la afirmación opuesta es verdad: en el espacio plano $\mathbb{R}^N$ (o el espacio de Minkowski) la conexión Γ^k_{ij} es identicamente cero en coordenadas cartesianas y por lo tanto el tensor de Riemann también es cero. Y dado el carácter tensorial de $R_{\mu\nu\rho}{}^{\lambda}$, si el tensor de Riemann es cero en coordenadas cartesianas también lo será en cualquier otro tipo de coordenadas curvilíneas, aunque las conexiones no necesariamente lo sean. (Es un ejercicio ilustrativo calcular $R_{\mu\nu\rho}{}^{\lambda}$ de $\mathbb{R}^2$ a partir de las componentes de la conexión (7.6) en coordenadas polares).

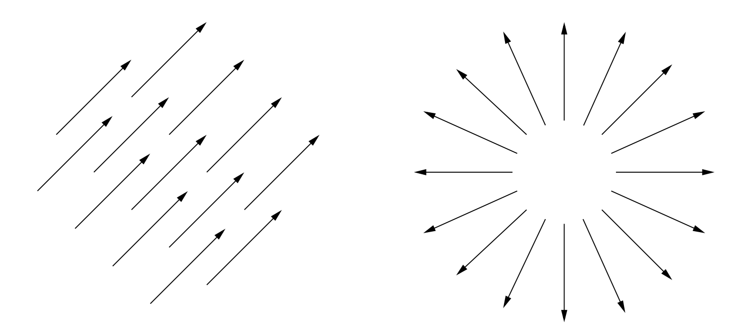


Figura 7.5: *Dos maneras distintas de definir el transporte paralelo en* $\mathbb{R}^2$*: la conexión de Levi-Civita define la idea intuitiva de paralelo (izquierda), pero la conexión inducida en el plano por la proyección estereográfica de una esfera implica llamar paralelos a los vectores de un campo radial (derecha), al provenir de vectores paralelos apuntando al Polo Norte de la esfera.*

Definimos por lo tanto que una *variedad es plana si todas las componentes del tensor de Riemann* $R_{\mu\nu\rho}{}^{\lambda}$ *son identicamente cero*. Por otro lado, una variedad tiene curvatura si por lo menos una componente de $R_{\mu\nu\rho}{}^{\lambda}$ es distinta de cero.

Merece la pena enfatisar una vez más el hecho que elecciones diferentes para la conexión llevan a diferentes expresiones para el tensor de Riemann, y por lo tanto a diferentes nociones de la curvatura en la misma variedad. Es un ejercicio instructivo ver que cambiando la conexión $\Gamma_{\mu\nu}^{\rho} \rightarrow \tilde{\Gamma}_{\mu\nu}^{\rho} = \Gamma_{\mu\nu}^{\rho} + K_{\mu\nu}^{\rho}$, con $K_{\mu\nu}^{\rho}$ un tensor arbitrario, el tensor de Riemann (y por lo tanto la curvatura) cambia como (ejerc.)

$$R_{\mu\nu\rho}{}^{\lambda} \;\rightarrow\; \tilde{R}_{\mu\nu\rho}{}^{\lambda} \;=\; R_{\mu\nu\rho}{}^{\lambda} \;+\; \nabla_\mu K_{\nu\rho}^{\lambda} \;-\; \nabla_\nu K_{\mu\rho}^{\lambda} \;+\; T_{\mu\nu}^{\sigma} K_{\sigma\rho}^{\lambda} \;+\; K_{\mu\sigma}^{\lambda} K_{\nu\rho}^{\sigma} \;-\; K_{\nu\sigma}^{\lambda} K_{\mu\rho}^{\sigma}. \quad (7.27)$$

Vemos por lo tanto que una variedad que tiene curvatura con cierta conexión $\Gamma_{\mu\nu}^{\rho}$, puede ser plana con una eleción apropidada de una nueva conexión $\tilde{\Gamma}_{\mu\nu}^{\rho}$, y vice versa.

7.4. Ejemplo concreto: distintas conexiones en $\mathbb{R}^2$

Ilustremos con un ejemplo concreto cómo las propiedades geométricas de una variedad pueden variar con las distintas elecciones de la conexión: consideremos $\mathbb{R}^2$ como variedad topológica, es decir, no inducimos ninguna métrica, ya que sólo miraremos el comportamiento de vectores bajo transporte paralelo. La noción intuitiva de lo que llamamos paralelo en $\mathbb{R}^2$ está definida por la conexión de Levi-Civita. Ya hemos dicho que en coordenadas polares (r, φ) ésta coge la forma (7.6)

$$\Gamma_{r\varphi}^{\varphi} \;=\; \Gamma_{\varphi r}^{\varphi} \;=\; r^{-1}, \qquad\qquad \Gamma_{\varphi\varphi}^{r} \;=\; -r, \quad (7.28)$$

y el tensor de Riemann, definida por esta conexión es idénticamente cero: $R_{\mu\nu\rho}{}^{\lambda} = 0$ (ejerc.). Esto coincide efectivamente con la intuición de que en el plano el resultado de transporte paralelo no depende de la curva seguida y que las distintas componentes de los vectores transportados no varían en una base cartesiana (véase Figura 7.5, izquierda).

Otra noción de paralelo, mucho menos intuitiva (pero también con cierta lógica) es la que induce la conexión

$$\tilde{\Gamma}_{rr}^{r} = \frac{-2r}{R_0^2 + r^2}, \qquad \tilde{\Gamma}_{\varphi r}^{\varphi} = \Gamma_{r\varphi}^{\varphi} = \frac{R_0^2 - r^2}{r(R_0^2 + r^2)}, \qquad \tilde{\Gamma}_{\varphi\varphi}^{r} = \frac{-r(R_0^2 - r^2)}{R_0^2 + r^2}. \quad (7.29)$$

Esta conexión no es plana, sino induce una curvatura no-trivial en el plano (ejerc.):

$$R_{r\varphi r}{}^{\varphi} = \frac{-4R_0^2}{(R_0^2+r^2)^2}, \qquad\qquad R_{r\varphi\varphi}{}^{r} = \frac{4R_0^2\, r^2}{(R_0^2+r^2)^2}. \tag{7.30}$$

En la sección 8.3 veremos que esta conexión proviene de una proyección estereográfica de la $\mathbb{S}^2$ a $\mathbb{R}^2$ y proyecta la noción de paralelismo en la esfera en el plano.[1] Un campo vectorial con simetría radial es por lo tanto invariante bajo transporte paralelo con la conexión (7.29), puesto que proviene de un campo vectorial constante en la esfera apuntando hacia el polo norte (véase Figura 7.5, derecha).

Dejamos como ejercicio demostrar que el tensor $K_{\mu\nu}^{\rho} = \tilde{\Gamma}_{\mu\nu}^{\rho} - \Gamma_{\mu\nu}^{\rho}$ con componentes

$$K_{rr}^{r} \;=\; K_{r\varphi}^{\varphi} \;=\; K_{\varphi r}^{\varphi} \;=\; \frac{-2r}{R_0^2+r^2}, \qquad\qquad K_{\varphi\varphi}^{r} \;=\; \frac{2r^3}{R_0^2+r^2}, \tag{7.31}$$

efectivamente relaciona las expresiones de los tensores de Riemannn de cada conexión según (7.27).

7.5. Tensores de curvatura

Habiendo definido el concepto de curvatura a través del tensor de Riemann $R_{\mu\nu\rho}{}^{\lambda}$, resulta útil estudiar también sus simetrías y algunas tensores formados de contracciones de $R_{\mu\nu\rho}{}^{\lambda}$. Para algunas simetrías será conveniente trabajar con la variante completamente covariante del tensor de Riemann, $R_{\mu\nu\rho\lambda} = g_{\lambda\sigma} R_{\mu\nu\rho}{}^{\sigma}$.

De la definición (7.23) está claro que el tensor de Riemann es antisimétrico en los primeros dos índices:

$$R_{\mu\nu\rho}{}^{\lambda} = -R_{\nu\mu\rho}{}^{\lambda}. \tag{7.32}$$

La identidad de Jacobi es una identidad básica en teoría de grupos, que determina el orden de las operaciones del álgebra. Aplicada en las derivadas covariantes toma la forma

$$[[\nabla_\mu, \nabla_\nu], \nabla_\rho] + [[\nabla_\nu, \nabla_\rho], \nabla_\mu] + [[\nabla_\rho, \nabla_\mu], \nabla_\nu] = 0. \tag{7.33}$$

De la identidad de Jacobi se puede derivar dos importantes identidades del tensor de Riemann, las llamadas identidades de Bianchi. La primera identidad de Bianchi se obtiene actuando con la identidad de Jacobi sobre un escalar ϕ. Tomando en cuenta (7.25), no es difícil ver que el tensor de Riemann y el tensor de torsión satisfacen la siguiente identidad

$$R_{\mu\nu\rho}{}^{\lambda} + R_{\nu\rho\mu}{}^{\lambda} + R_{\rho\mu\nu}{}^{\lambda} - \nabla_\mu T_{\nu\rho}^{\lambda} - \nabla_\nu T_{\rho\mu}^{\lambda} - \nabla_\rho T_{\mu\nu}^{\lambda} + T_{\mu\nu}^{\sigma} T_{\rho\sigma}^{\lambda} + T_{\nu\rho}^{\sigma} T_{\mu\sigma}^{\lambda} + T_{\rho\mu}^{\sigma} T_{\nu\sigma}^{\lambda} = 0. \tag{7.34}$$

La segunda identidad de Bianchi se obtiene actuando con la identidad de Jacobi sobre un vector contravariante V^{λ},

$$\nabla_\mu R_{\nu\rho\sigma}{}^{\lambda} + \nabla_\nu R_{\rho\mu\sigma}{}^{\lambda} + \nabla_\rho R_{\mu\nu\sigma}{}^{\lambda} - T_{\mu\nu}^{\tau} R_{\rho\tau\sigma}{}^{\lambda} - T_{\nu\rho}^{\tau} R_{\mu\tau\sigma}{}^{\lambda} - T_{\rho\mu}^{\tau} R_{\nu\tau\sigma}{}^{\lambda} = 0. \tag{7.35}$$

Contrayendo el segundo índice del tensor de Riemann con el cuarto, se obtiene el *tensor de Ricci*:

$$R_{\mu\nu} \;=\; R_{\mu\rho\nu}{}^{\rho} \;=\; g^{\rho\lambda} R_{\mu\rho\nu\lambda}. \tag{7.36}$$

Por lo tanto, en función de la conexión viene dado por

$$R_{\mu\nu} = \partial_\mu \Gamma_{\lambda\nu}^{\lambda} - \partial_\lambda \Gamma_{\mu\nu}^{\lambda} + \Gamma_{\mu\sigma}^{\lambda} \Gamma_{\lambda\nu}^{\sigma} - \Gamma_{\mu\nu}^{\lambda} \Gamma_{\sigma\lambda}^{\sigma}. \tag{7.37}$$

[1] De hecho, en el siguiente capítulo veremos que las expresiones (7.30) del tensor de Riemann corresponden a una esfera bidimensional. El plano topológico equipada con la conexión (7.29) es geométricamente una esfera.

En general en tensor de Ricci no es simétrico, puesto que la parte antisimétrica toma la forma (ejerc.)

$$R_{\mu\nu} - R_{\nu\mu} = R_{\mu\nu\rho}{}^{\rho} - \nabla_\rho T^\rho_{\mu\nu} - \nabla_\mu T^\rho_{\nu\rho} + \nabla_\nu T^\rho_{\mu\rho} - T^\lambda_{\mu\nu} T^\rho_{\lambda\rho} \tag{7.38}$$

Veremos más adelante que las fórmulas (7.34), (7.35) y (7.38) se simplificarán mucho si trabajamos con la conexión de Levi-Civita.

Contrayendo los dos índices del tensor de Ricci, se obtiene el *escalar de Ricci* o el *escalar de curvatura*:

$$R = g^{\mu\nu} R_{\mu\nu}. \tag{7.39}$$

Como dice el nombre, el escalar de Ricci es un escalar y por lo tanto un invariante bajo cambios generales de coordenadas.

Finalmente, un tensor que juega un papel importante en la relatividad general es el *tensor de Einstein*, construido del tensor de Ricci, el escalar de Ricci y la métrica como

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R. \tag{7.40}$$

Igual que el tensor de Ricci, el tensor de Einstein en general no es simétrico. Por otro lado contrayendo la segunda identidad de Bianchi (7.35) con $\delta^\rho_\lambda g^{\mu\sigma}$ obtenemos la identidad

$$2\nabla_\mu G_\nu{}^\mu \;+\; 2T^\rho_{\mu\nu} R_\rho{}^\mu \;+\; T^\lambda_{\mu\rho} R_{\nu\lambda}{}^{\mu\rho} \;-\; \nabla_\rho g^{\mu\lambda}\Big[\delta^\rho_\mu R_{\nu\lambda} \;+\; \delta^\rho_\nu R_{\mu\lambda} \;+\; R_{\mu\nu\lambda}{}^\rho\Big] = 0. \tag{7.41}$$

También esta identidad simplifica mucho con la elección de Levi-Civita y jugará un papel importante a la hora de construir las ecuaciones del campo gravitatorio.

7.6. Geodésicas afines y métricas

En $\mathbb{R}^N$ las rectas son curvas especiales por dos razones distintas. La primera razón es que una recta es la curva más corta entre dos puntos p y q. La segunda es que es la única curva donde el vector tangente está transportado paralelamente a si mismo a lo largo de la curva. En esta sección veremos cómo podemos generalizar cada uno de estos conceptos a variedades más generales.

Una primera observación es que en principio las dos propiedades de las rectas en $\mathbb{R}^N$ no tienen nada que ver el uno con el otro. Una habla de la distancia entre dos puntos, mientras que la otra de paralelismo, y necesitamos dos conceptos matemáticas distintos para definirlos: la métrica para distancias y ángulos y la conexión afín para definir lo que entendemos por paralelo.

No es de extrañar entonces que podemos generalizar el concepto de la recta en $\mathbb{R}^N$ de dos maneras distintas a espacios arbitrarios: una, a través de la métrica, como la curva que minimiza la distancia entre dos puntos y la otra, a través de la conexión, como la curva cuyo vector tangente es (covariantemente) constante. En general, para conexiones arbitrarias estas dos generalizaciones resultarán en curvas distintas.

Consideramos primero la generalización de la segunda característica: una *geodésica afín* es aquella curva $x^\mu(\tau)$, cuyo vector tangente $u^\mu = \dot{x}^\mu(\tau) = dx^\mu/d\tau$ esté transportado paralelamente a si mismo a lo largo de la curva. En otras palabras, las geodésicas afines son las curvas más "rectas" que podemos definir en una variedad arbitraria. Con (7.21) vemos entonces que las geodésicas afines están definidas por la ecuación

$$u^\nu \nabla_\nu u^\mu \;=\; 0 \tag{7.42}$$

Usando que $u^\nu \partial_\nu = d/d\tau$ es la derivada (parcial) direccional en la dirección de la curva, podemos escribir esta ecuación como una ecuación diferencial para $x^\mu(\tau)$:

$$\ddot{x}^\mu \;+\; \Gamma^\mu_{\nu\rho} \dot{x}^\nu \dot{x}^\rho \;=\; 0. \tag{7.43}$$

Obsérvese que en esta ecuación τ es simplemente el parámetro afín que parametriza la curva $x^\mu(\tau)$ y no (necesariamente) el tiempo propio en el caso de curvas temporales. Es más, hasta que no introduzcamos la métrica, no tenemos manera de calcular la longitud de las geodésicas afines.

Obsérvese también que en el último término de (7.43) $\dot{x}^\nu$ y $\dot{x}^\rho$ aparecen de manera simétrica, de modo que sólo la parte simétrica de la conexión contribuye a la determinación de las geodésicas afines. En otras palabras, dos conexiónes $\Gamma_{\mu\nu}^\rho$ y $\tilde{\Gamma}_{\mu\nu}^\rho$ que difieren por su torsión, $\tilde{\Gamma}_{\mu\nu}^\rho = \Gamma_{\mu\nu}^\rho + T_{\mu\nu}^\rho$, tienen las mismas geodésicas afines.

La ecuación (7.43) no debería ser del todo desconocida: para $\mathbb{R}^N$ equipada con la conexión de Levi-Civita (la noción intuitiva de lo que llamamos paralelo), los coeficientes de conexión son cero en coordenadas cartesianas, $\Gamma_{ij}^k = 0$, y la ecuación (7.43) se reduce a la conocida ecuación de la recta, $\ddot{x}^i(\tau) = 0$. Además, es ilustrativo comprobar que al aplicar un cambio general de coordenadas $x^i = x^i(y^\mu)$ a la ecuación de la recta y usando la regla de transformación de la conexión (7.15), se obtiene precisamente la forma general (7.43) de la ecuación de la geodésica (ejerc.).

La otra generalización de una recta en variedades arbitrarios es buscando la curva más corta entre dos puntos.[2] Las geodésicas de este tipo se llaman *geodésicas métricas* y, como hemos dicho, para conexiones arbitrarias no coinciden con las geodésicas afines.

La mejor manera de definir la distancia más corta entre dos puntos es a través de cálculo variacional: una *geodésica métrica* es la curva $x^\mu(\tau)$ entre dos puntos p y q cuya longitud

$$s = \int_p^q d\tau \sqrt{g_{\mu\nu}(x(\tau))\,\dot{x}^\mu(\tau)\,\dot{x}^\nu(\tau)} \tag{7.44}$$

es estacionaria bajo pequeñas variaciones de la curva y de la métrica,

$$x^\mu(\tau) \longrightarrow x^{\,\mu}(\tau) + \delta x^\mu(\tau), \qquad\qquad g_{\mu\nu}(x) \longrightarrow g_{\mu\nu}(x) + \partial_\rho g_{\mu\nu}(x)\,\delta x^\rho. \tag{7.45}$$

La geodésica corresponde por lo tanto a la curva para la cual $\delta s = 0$ y ésta viene dada por la ecuación de Euler-Lagrange,

$$\frac{d}{d\tau}\Big(\frac{\delta L}{\delta \dot{x}^\mu}\Big) - \frac{\delta L}{\delta x^\mu} = 0, \tag{7.46}$$

donde $L = ds/d\tau = \sqrt{g_{\mu\nu}\,\dot{x}^\mu\,\dot{x}^\nu}$ es la funcional que aparece en (7.44).

La manera más fácil de proceder es multiplicar (7.46) con un factor $2L$, para eliminar la raiz cuadrada. La ecuación de Euler-Lagrange se escribe entonces como

$$\frac{d}{d\tau}\Big(\frac{\delta L^2}{\delta \dot{x}^\mu}\Big) - \frac{\delta L^2}{\delta x^\mu} = 2\frac{\delta L}{\delta \dot{x}^\mu}\frac{dL}{d\tau}, \tag{7.47}$$

y rellenando la expresión para L obtenemos para el lado izquierdo

$$\begin{aligned}
\frac{d}{d\tau}\Big(\frac{\delta L^2}{\delta \dot{x}^\mu}\Big) - \frac{\delta L^2}{\delta x^\mu} &= \frac{d}{d\tau}\Big(2g_{\mu\nu}\,\dot{x}^\nu\Big) - \partial_\mu g_{\nu\rho}\,\dot{x}^\nu\dot{x}^\rho \\
&= 2g_{\mu\nu}\,\ddot{x}^\nu + 2\partial_\rho g_{\mu\nu}\,\dot{x}^\nu\,\dot{x}^\rho - \partial_\mu g_{\nu\rho}\,\dot{x}^\nu\,\dot{x}^\rho \\
&= 2g_{\mu\nu}\,\ddot{x}^\nu + \Big[\partial_\nu g_{\mu\rho} + \partial_\rho g_{\nu\mu} - \partial_\mu g_{\nu\rho}\Big]\,\dot{x}^\nu\,\dot{x}^\rho \\
&= 2g_{\mu\lambda}\,\ddot{x}^\lambda + 2g_{\mu\lambda}\left\{{}^{\lambda}_{\nu\rho}\right\}\dot{x}^\nu\dot{x}^\rho,
\end{aligned} \tag{7.48}$$

[2]La definición de la geodésica como la distancia más corta entre dos punto es solamente válida en espacios riemannianos y en espacios lorentzianos en el caso de geodésicas espaciales. Geodésicas temporales *maximizan* el tiempo propio, debido al signo relativo entre las coordenadas.

donde en la segunda igualdad hemos usado que $dg_{\mu\nu}(x(\tau))/d\tau = \partial_\rho g_{\mu\nu}\dot{x}^\rho$ y en la última igualdad hemos definido los *símbolos de Christoffel*

$$\left\{ {}^{\rho}_{\mu\nu} \right\} \;=\; \tfrac{1}{2}\, g^{\rho\lambda} \Big[\partial_\mu g_{\lambda\nu} \,+\, \partial_\nu g_{\mu\lambda} \,-\, \partial_\lambda g_{\mu\nu}\Big]. \tag{7.49}$$

Por otro lado, el lado derecho de (7.47) da

$$2\,\frac{\delta L}{\delta \dot{x}^\mu}\frac{dL}{d\tau} \;=\; 2\,\frac{g_{\mu\nu}\,\dot{x}^\nu}{\sqrt{g_{\mu\nu}\,\dot{x}^\mu\,x^\nu}}\,\frac{d}{d\tau}\Big(\frac{ds}{d\tau}\Big) \;=\; 2L^{-1}\,g_{\mu\nu}\,\dot{x}^\nu\,\frac{d^2s}{d\tau^2} \;=\; 2g_{\mu\lambda}\,\dot{x}^\lambda\,\left(\frac{\ddot{s}}{\dot{s}}\right). \tag{7.50}$$

Sustituyendo los dos lados en (7.47) y multiplicando con $g^{\sigma\mu}$, vemos que la ecuación de Euler-Lagrange se reduce a

$$\ddot{x}^\sigma \,+\, \left\{ {}^{\sigma}_{\nu\rho} \right\} \dot{x}^\nu \dot{x}^\rho \;=\; \dot{x}^\sigma\,\left(\frac{\ddot{s}}{\dot{s}}\right). \tag{7.51}$$

Nótese que el lado izquierdo de esta ecuación contiene sólo cantidades geométricas (las coordenadas de la curva y las componentes de la métrica), mientras que el lado derecho depende de la parametrización particular de la curva. Esta última no tiene significado físico (geométrico) y siempre podemos elegir una parametrización lineal de la curva, $s(\tau) = \tau + a$, de modo que el lado derecho es cero y la ecuación (7.51) se reduce a

$$\ddot{x}^\sigma \,+\, \left\{ {}^{\sigma}_{\nu\rho} \right\} \dot{x}^\nu\,\dot{x}^\rho \;=\; 0. \tag{7.52}$$

Obviamente se puede usar cualquier otra parametrización, siempre y cuando se usa (7.51) en ese caso. Como la parametrización de la curva no tiene significado intrínsico, las dos ecuaciones (7.51) y (7.52) describen la misma curva en la variedad.

Finalmente, para métricas lorentzianas existen tres tipos de geodésicas: temporales, espaciales y nulas, dependiento si la distancia entre los puntos p y q es temporal, espacial o nula. En el caso de que la geodésica es temporal, el parámetro τ tiene la interpretación del tiempo propio de la partícula que viaja a lo largo de la geodésica. Si la geodésica es espacial o nula, el parámetro τ no tiene un significado físico especial. Para distinguir los tres casos hay que añadir a (7.52) una ecuación más que especifica el tipo de geodésica a través del vector tangente a la curva:

$$g_{\mu\nu}\,\dot{x}^\mu\,\dot{x}^\nu \;=\; \varepsilon, \qquad \text{donde} \qquad \varepsilon = \begin{cases} \;\;1 & \text{si } x^\mu(\tau) \text{ es temporal,} \\ \;\;0 & \text{si } x^\mu(\tau) \text{ es nulo,} \\ -1 & \text{si } x^\mu(\tau) \text{ es espacial.} \end{cases} \tag{7.53}$$

En el siguiente capítulo veremos como la diferencia entre geodésicas afines y métricas desaparece al usar la conexión de Levi-Civita.

Capítulo 8

Cálculo tensorial con la conexión de Levi-Civita

No te preocupes por tus problemas con las matemáticas; te aseguro que los mios son mucho mayores.
(A. Einstein, 1943, en una carta a una niña de 9 años)

En la capítulo anterior hemos insistido mucho en que las propiedades geométricas de una variedad dependen mucho de la conexión que uno elija y que ésta es en principio arbitraria. Sin embargo, en este capítulo veremos que en una variedad equipada con una métrica, existe una conexión especial, llamada la conexión métrica o la de Levi-Civita, que simplifica muchas de las ecuaciones geométricas. La conexión de Levi-Civita resultará también ser importante desde el punto de vista físico, puesto que es ésta la que se aparece de manera natural en la relatividad general.

8.1. La conexión de Levi-Civita

Ya hemos mencionado en anteriormente que en una variedad equipada con una conexión arbitraria $\Gamma^{\rho}_{\mu\nu}$, la métrica $g_{\mu\nu}$ en general no tiene nada que ver con la conexión. Sin embargo se puede demostrar que en cualquier variedad existe siempre una conexión única, la *conexión de Levi-Civita*, que tiene una relación especial con la métrica. En particular, resulta que la conexión de Levi-Civita, y por lo tanto también los tensores de curvatura, están completamente determinadas en términos de la métrica.

La conexión de Levi-Civita satisface dos condiciones:

1. La conexión es simétrica: $\Gamma^{\rho}_{\mu\nu} = \Gamma^{\rho}_{\nu\mu}$.
2. La derivada covariante de la métrica es cero: $\nabla_\mu g_{\nu\rho} = 0$.

La primera condición implica que el tensor de torsión $T^{\rho}_{\mu\nu}$ es cero y por lo tanto el conmutador de dos derivadas covariantes viene dado sólo por el tensor de Riemann:

$$[\nabla_\mu, \nabla_\nu] V^\lambda = R_{\mu\nu\rho}{}^{\lambda} V^\rho. \tag{8.1}$$

La segunda condición se llama la *compatibilidad con la métrica*. Las conexiónes que son compatibles con la métrica tienen varias propiedades que simplifican mucho las propiedades geométricas de la variedad.

Primero la compatibilidad de la métrica implica que el producto escalar entre dos vectores es invariante bajo transporte paralelo de los vectores. Si transportamos dos vectores V^μ y W^μ de manera paralela lo largo de una curva $x^\mu(\tau)$,

$$u^\mu \nabla_\mu V^\nu \;=\; 0 \;=\; u^\mu \nabla_\mu W^\nu, \tag{8.2}$$

con $u^\mu = \dot{x}^\mu$ el vector tangente a la curva, entonces el producto escalar $g_{\mu\nu} V^\mu W^\nu$ cambia como resultado del transporta paralelo como

$$\begin{aligned} u^\mu \nabla_\mu (g_{\nu\rho} V^\nu W^\rho) &= u^\mu \nabla_\mu g_{\nu\rho} \, V^\nu \, W^\rho \;+\; g_{\nu\rho} \, u^\mu \nabla_\mu V^\nu \, W^\rho \;+\; g_{\nu\rho} \, V^\nu \, u^\mu \nabla_\mu W^\rho \\ &= u^\mu \nabla_\mu g_{\nu\rho} \, V^\nu \, W^\rho, \end{aligned} \tag{8.3}$$

donde en la segunda igualdad hemos usado que bajo transporte paralelo a lo largo de la curva V^μ y W^μ cumplen (8.2). Vemos por lo tanto que el producto escalar es invariante, es decir $u^\mu \nabla_\mu (V_\rho W^\rho) = 0$, para cualquier curva $x^\mu(\tau)$ y cualesquiera dos vectores V^μ y W^μ, sólo si la conexión es compatible con la métrica. Nótese que la condición de compatibilidad con la métrica es una condición más bien física que matemática, puesto que implica que la norma de un vector no cambia al hacer transporte paralelo a lo largo de una curva, algo que se espera por razones físicas.

La segunda ventaja de la compatibilidad con la métrica es que implica que la derivada covariante conmuta con subir, bajar o contraer índices:

$$g_{\mu\nu} \nabla_\rho S^{\mu\nu} \;=\; \nabla_\rho (g_{\mu\nu} S^{\mu\nu}) \;=\; \nabla_\rho S^\mu{}_\mu. \tag{8.4}$$

Esta propiedad no es verdad para derivadas covariantes con conexiones que no sean compatibles con la métrica.

Dadas estas dos condiciones, la conexión queda completamente determinada por la métrica. Considérese las tres derivadas covariantes de la métrica:

$$\begin{aligned} 0 &= \nabla_\mu g_{\nu\rho} \;=\; \partial_\mu g_{\nu\rho} \;-\; \Gamma^\lambda_{\mu\nu} \, g_{\lambda\rho} \;-\; \Gamma^\lambda_{\mu\rho} \, g_{\nu\lambda}, \\ 0 &= \nabla_\nu g_{\rho\mu} \;=\; \partial_\nu g_{\rho\mu} \;-\; \Gamma^\lambda_{\nu\rho} \, g_{\lambda\mu} \;-\; \Gamma^\lambda_{\nu\mu} \, g_{\rho\lambda}, \\ 0 &= \nabla_\rho g_{\mu\nu} \;=\; \partial_\rho g_{\mu\nu} \;-\; \Gamma^\lambda_{\rho\mu} \, g_{\lambda\nu} \;-\; \Gamma^\lambda_{\rho\nu} \, g_{\mu\lambda}. \end{aligned} \tag{8.5}$$

Si sumamos las primeras dos expresiones y le restamos la tercera, utilizando el hecho de que tanto la métrica como la conexión sean simétricas, obtenemos

$$\partial_\mu g_{\nu\rho} \;+\; \partial_\nu g_{\rho\mu} \;-\; \partial_\rho g_{\mu\nu} - 2\Gamma^\lambda_{\mu\nu} \, g_{\lambda\rho} \;=\; 0, \tag{8.6}$$

o, multiplicando con $g^{\sigma\rho}$,

$$\Gamma^\sigma_{\mu\nu} \;=\; \frac{1}{2} \, g^{\sigma\lambda} \Big(\partial_\mu g_{\lambda\nu} \;+\; \partial_\nu g_{\mu\lambda} \;-\; \partial_\lambda g_{\mu\nu} \Big). \tag{8.7}$$

Podemos por lo tanto expresar la conexión de Levi-Civita completamente en función de las componentes de la métrica y sus derivadas.

Vemos entonces que en una variedad equipada con la conexión de Levi-Civita, todas las propiedades geométricas están determinadas únicamente por la métrica: con ella se puede determinar las distancias entre distintos puntos y los angulos entre vectores, pero además al determinar completamente la conexión de Levi-Civita, los conceptos de transporte paralelo y curvatura, en particular el tensor de Riemann, quedan definidos en términos de la métrica.

El hecho de que la conexión de Levi-Civita sea la única conexión que satisface (simultaneamente) las condiciones de simetría y compatibilidad con la métrica, hace que esta conexión juegue un papel preferido dentro de las matemáticas y la física. En particular, salvo escasas excepciones se suele utilizar esta conexión en la relatividad general. Además resulta que la noción de transporte paralelo que induce la conexión de Levi-Civita es la que más se aproxima a nuestra intuición de paralelo en espacios curvos.

8.2. Las simetrías de los tensores de curvatura

Estará claro que las dos condiciones que determinan la conexión de Levi-Civita simplifican mucho las propiedades geométricas del espacio: La simetría de la conexión hace que la torsión $T^{\rho}_{\mu\nu}$ desaparece de todas las expresiones y la compatibilidad con la métrica) permite subir y bajar índices detrás de derivadas covariantes con mucha facilidad.

Además, la conexión de Levi-Civita hace que los tensores de curvatura tienen más simetrías que en el caso general y satisfacen relaciones más sencillas. Ya hemos visto en (7.32) que el tensor de Riemann es antisimétrico en los primeros dos índices, pero resulta que $R_{\mu\nu\rho\lambda}$ también es antisimétrico en las últimas dos índices, si la conexión es compatible con la métrica

$$R_{\mu\nu\rho\lambda} = -R_{\mu\nu\lambda\rho}. \tag{8.8}$$

Esta última propiedad se demuestra fácilmente, exigiendo que el conmutador de dos derivadas covariantes actuando sobre la métrica sea cero, $[\nabla_\mu, \nabla_\nu]g_{\rho\lambda} = 0$ (ejerc.). Por otro lado, las identidades de Bianchi (7.34) y (7.35) se reducen con la conexión de Levi-Civita a

$$\begin{aligned} R_{\mu\nu\rho\lambda} \ + \ R_{\nu\rho\mu\lambda} \ + \ R_{\rho\mu\nu\lambda} \ &= \ 0, \\ \nabla_\mu R_{\nu\rho\sigma\lambda} \ + \ \nabla_\nu R_{\rho\mu\sigma\lambda} \ + \ \nabla_\rho R_{\mu\nu\sigma\lambda} \ &= \ 0. \end{aligned} \tag{8.9}$$

Finalmente, las reglas de antisimetría (7.32) y (8.8), en combinación con la primera identidad de Bianchi (8.9), implican que el tensor de Riemann es simétrico bajo el intercambio del primer par de índices con el segundo par (ejerc.),

$$R_{\mu\nu\rho\lambda} \ = \ R_{\rho\lambda\mu\nu}. \tag{8.10}$$

Estas condiciones reducen el número de grados de libertad del tensor de Riemann, es decir el número de componentes independientes. Un pequeño cálculo combinatorio muestra que el tensor de Riemann en N dimensiones tiene $\frac{1}{12}N^2(N^2-1)$ grados de libertad, lo que en $N = 4$ corresponde a 20.

Otra ventaja de la conexión de Levi-Civita es el tensor de Ricci $R_{\mu\nu}$ es ahora la única contracción independiente de Riemann. Donde con una conexión arbitraria los tensores

$$R_{\mu\nu} = R_{\mu\lambda\nu}{}^{\lambda}, \qquad \tilde{R}_{\mu\nu} = g^{\rho\lambda}R_{\mu\rho\lambda\nu}, \qquad \bar{R}_{\mu\nu} = R_{\mu\nu\lambda}{}^{\lambda}, \tag{8.11}$$

son en principio contracciones independientes, las simetrías del tensor de Riemann hacen ahora que estas contracciones sean o bien idénticamente cero, o bien proporcionales al tensor de Ricci.

También está claro de (7.38) que con Levi-Civita tanto el tensor de Ricci $R_{\mu\nu}$ como el tensor de Einstein $G_{\mu\nu}$ son simétricos en los dos índices. Fíjese que los términos de torsión son cero por la simetría de la conexión, mientras el término $R_{\mu\nu\rho}{}^{\rho}$ es cero debido a la propiedad (8.8), es decir, debido a la compatibilidad de la conexión con al métrica. Finalmente, una de las propiedades importantes del tensor de Einstein para la relatividad general es que su divergencia es cero,

$$\nabla_\mu G^{\mu\nu} \ = \ 0, \tag{8.12}$$

como se ve fácilmente aplicando las condiciones de la conexión de Levi-Civita a la ecuación (7.41). Esto impone la siguiente relación entre las derivadas del tensor y el escalar de Ricci:

$$\nabla_\mu R_\nu{}^\mu = \frac{1}{2}\partial_\nu R. \tag{8.13}$$

La propiedad (8.12) jugará en importante papel en la construcción de las ecuaciones del campo gravitacional, como veremos en el Capítulo 10.

8.3. Ejemplo concreto: tensores de curvatura de la $\mathbb{S}^2$

Como ejemplo concreto, repasaremos la esfera bidimensional que hemos encontrado en el Capítulo 6. La métrica en coordenadas azimutales era

$$ds^2 \;=\; R_0^2 \Big(d\theta^2 \;+\; \sin^2\theta d\varphi^2\Big), \tag{8.14}$$

donde R_0 era el radio de la esfera. A través de la fórmula (8.7) no es difícil ver que los símbolos de Christoffel no nulos son (ejerc.)

$$\Gamma^{\varphi}_{\theta\varphi} = \Gamma^{\varphi}_{\varphi\theta} = \operatorname{cotg}\theta, \qquad\qquad \Gamma^{\theta}_{\varphi\varphi} = -\sin\theta\cos\theta, \tag{8.15}$$

y por lo tanto las componentes no-triviales del tensor de Riemann vienen dadas por (ejerc.)

$$R_{\theta\varphi\theta}{}^{\varphi} = -R_{\varphi\theta\theta}{}^{\varphi} = -1, \qquad\qquad R_{\theta\varphi\varphi}{}^{\theta} = -R_{\varphi\theta\varphi}{}^{\theta} = \sin^2\theta. \tag{8.16}$$

Según la fórmula de la sección 8.1, el tensor de Riemann completamente covariante tiene en dos dimensiones sólo una componente independiente. Efectivamente, de (8.16) se ve fácilmente que (ejerc.)

$$R_{\theta\varphi\theta\varphi} = -R_0^2\sin^2\theta, \tag{8.17}$$

y que todas las demás componentes o son cero, o están relacionadas con esta a través de las fórmulas (7.32), (8.8) y (8.10). El tensor de Ricci se puede obtener o bien a través de la contracción del tensor de Riemann, o bien a través de la fórmula (7.37). Las componentes no-triviales son (ejerc.)

$$R_{\theta\theta} = -1, \qquad\qquad R_{\varphi\varphi} = -\sin^2\theta, \tag{8.18}$$

y por lo tanto el escalar de Ricci viene dado por

$$R = -2R_0^{-2}. \tag{8.19}$$

La esfera bidimensional (igual que las demás esferas en otras dimensiones) es un espacio con mucha simetría y esto se refleja en sus tensores de curvatura. Obsérvese que las componentes del tensor de Riemann satisfacen la expresión

$$R_{\mu\nu\rho\lambda} = R_0^{-2}\Big(g_{\mu\lambda}g_{\nu\rho} - g_{\mu\rho}g_{\nu\lambda}\Big). \tag{8.20}$$

En el Capítulo 13 veremos que esta es la definición de los espacios con curvatura constante, o, equivalentemente, los espacios máximamente simétricos. Contrayendo (8.20) se obtiene una expresión parecida para el tensor de Ricci,

$$R_{\mu\nu} = -R_0^{-2}g_{\mu\nu}, \tag{8.21}$$

que efectivamente es una condición que satisfacen las componentes encontrados en (8.18).

Es instructivo hacer el mismo cálculo utilizando en la esfera las coordenadas estereográficas (6.45),

$$ds^2 \;=\; \frac{4R_0^4}{(R_0^2+r^2)^2}\,\Big[dr^2 \;+\; r^2 d\varphi^2\Big]. \tag{8.22}$$

En estas coordenadas los símbolos de Christoffel vienen dados por (ejerc.)

$$\Gamma^{r}_{rr} = \frac{-2r}{R_0^2+r^2}, \qquad \Gamma^{\varphi}_{\varphi r} = \Gamma^{\varphi}_{r\varphi} = \frac{R_0^2-r^2}{r(R_0^2+r^2)}, \qquad \Gamma^{r}_{\varphi\varphi} = \frac{-r(R_0^2-r^2)}{R_0^2+r^2}. \tag{8.23}$$

Obsérvese que estas expresiones son muy diferentes tanto de las expresiones (7.6) para $\mathbb{R}^2$ en coordenadas polares, como (8.15 para la $\mathbb{S}^2$ en coordenadas estándar respectivamente. En particular vemos que en estas coordenadas hay una componente no-trivial más que en los otros dos casos, debido a que los símbolos de Christoffel no transforman como un tensor.

Utilizando estas expresiones, no es muy difícil calcular los tensores de curvatura. Para el tensor de Riemann tenemos que[1]

$$R_{r\varphi r}{}^{\varphi} = \frac{-4R_0^2}{(R_0^2+r^2)^2}, \qquad R_{r\varphi\varphi}{}^{r} = \frac{4R_0^2\, r^2}{(R_0^2+r^2)^2}, \qquad R_{r\varphi r\varphi} = \frac{-16R_0^6\, r^2}{(R_0^2+r^2)^4}, \tag{8.24}$$

mientras que el tensor y el escalar de Ricci son de la forma

$$R_{rr} = \frac{-4R_0}{(R_0^2+r^2)^2}, \qquad R_{\varphi\varphi} = \frac{-4R_0\, r^2}{(R_0^2+r^2)^2}, \qquad R = -2R_0^2. \tag{8.25}$$

Nótese que también en estas coordenadas los tensores de curvatura satisfacen las identidades (8.20) y (8.21). Lógico, puesto que (8.20) y (8.21) son expresiones covariantes: si son válidas en un sistema de referencia, también lo son en todos los demás.

Por último, se podrían haber sacado las expresiones anteriores directamente a través de las reglas de transformación de los tensores. Dada la relación (6.43) entre θ y r, tenemos que

$$\frac{\partial\theta}{\partial r} = \frac{-2R_0}{R_0^2+r^2}, \tag{8.26}$$

y por lo tanto los tensores de curvatura transforman como

$$\begin{aligned} R_{rr} &= \frac{\partial\theta}{\partial r}\frac{\partial\theta}{\partial r}R_{\theta\theta} = \frac{4R_0^2}{(R_0^2+r^2)^2}(-1) = \frac{-4R_0^2}{(R_0^2+r^2)^2}, \\ R_{\varphi\varphi} &= -\sin\theta = \frac{-4R_0\, r^2}{(R_0^2+r^2)^2}, \\ R_{r\varphi r\varphi} &= \frac{\partial\theta}{\partial r}\frac{\partial\theta}{\partial r}R_{\theta\varphi\theta\varphi} = \frac{4R_0^2}{(R_0^2+r^2)^2}(-R_0^2\sin^2\theta) = \frac{-16R_0^6\, r^2}{(R_0^2+r^2)^4}. \end{aligned} \tag{8.27}$$

8.4. Operadores diferenciales en variedades arbitrarias

Ya que hemos construido la derivada covariante ∇_μ como el operador que transforma como un vector bajo cambios generales de coordenadas, tenemos que redefinir los operadores diferenciales conocidos de $\mathbb{R}^3$, encontrados en el Capítulo 4, tal que también sean válidas en variedades arbitrarias.

Dado que la derivada covariante actuando de un escalar coincide con la derivada parcial, $\nabla_\mu\phi = \partial_\mu\phi$, la definición del gradiente de una función escalar no sufre cambios con respecto a la definición en $\mathbb{R}^3$.

Sin embargo, para que la divergencia, el rotacional y el laplaciano transformen bien bajo cambios generales de coordenadas, es preciso construir estos operadores diferenciales con derivadas covariantes:

$$\vec{\nabla}\cdot\vec{A} = \nabla_\mu A^\mu, \qquad (\vec{\nabla}\times\vec{A})^\mu = \varepsilon^{\mu\nu\rho}\nabla_\nu A_\rho, \qquad \nabla^2\phi = \nabla_\mu\partial^\mu\phi. \tag{8.28}$$

Estas definiciones, aunque escritas de esta formas son válidas para conexiones arbitrarios, simplifican bastante si la conexión es la de Levi-Civita. En este caso se puede escribir el rotacional

[1] Nótese que hemos encontrado estas expresiónes antes, en (7.29) y (7.30), cuando estudiabamos las distintas definiciones de paralelismo en $\mathbb{R}^2$. Vemos por lo tanto lo que ya anunciábamos en la sección 7.4: que la conexión (7.29) no es la conexión plana de $\mathbb{R}^2$, sino la conexión de Levi-Civita de la $\mathbb{S}^2$ en coordenadas (8.22). En la sección 7.4 la podiamos asociar a la topología del plano $\mathbb{R}^2$, porque no estabamos considerando las porpiedades geométricas. Como dijimos, geométricamente hablando, la métrica (8.22) representa una esfera.

puramente en términos de derivadas parciales, ya que la contracción de la conexión simétrica y el tensor de Levi-Civita antisimétrico es idénticamente cero:

$$(\vec{\nabla} \times \vec{A})^\mu \;=\; \varepsilon^{\mu\nu\rho} \Big(\partial_\nu A_\rho - \Gamma_{\nu\rho}^\lambda A_\lambda\Big) \;=\; \varepsilon^{\mu\nu\rho}\,\partial_\nu A_\rho. \tag{8.29}$$

También las expresiones para la divergencia y el laplaciano son más sencillas. De la misma definición (8.7) de la conexión de Levi-Civita se puede demostrar con el uso de (A.23) que el símbolo de Christoffel con dos índices contraídos se puede escribir como (ejerc.)

$$\Gamma_{\mu\nu}^\nu \;=\; \frac{1}{\sqrt{|g|}}\partial_\mu\sqrt{|g|}. \tag{8.30}$$

Usando esta propiedad, la divergencia de un vector y el laplaciano de un escalar toman la forma

$$\vec{\nabla}\cdot\vec{A} \;=\; \frac{1}{\sqrt{|g|}}\partial_\mu\Big[\sqrt{|g|}\,A^\mu\Big], \qquad \nabla^2\phi \;=\; \frac{1}{\sqrt{|g|}}\partial_\mu\Big[\sqrt{|g|}\,\partial^\mu\phi\Big]. \tag{8.31}$$

No es difícil ver que esta expresión para la divergencia es incluso válida para tensores antisimétricos: si $T^{\mu_1\cdots\mu_m}$ es un tensor completamente antisimétrico, entonces (ejerc.)

$$\nabla_\mu T^{\mu\nu_1\ldots\nu_{m-1}} \;=\; \frac{1}{\sqrt{|g|}}\partial_\mu\Big[\sqrt{|g|}\,T^{\mu\nu_1\ldots\nu_{m-1}}\Big]. \tag{8.32}$$

Es instructivo ver hasta qué punto las conocidas identidades de las diferentes operadores diferenciales actuando unos sobre otros generalizan de $\mathbb{R}^N$ ($\mathbb{R}^3$ en caso de rotacionales) a variedades arbitrarias. Claramente, la divergencia de un gradiente sigue siendo el laplaciano,

$$\vec{\nabla}\cdot(\vec{\nabla}\phi) \;=\; \nabla_\mu\partial^\mu\phi \;=\; \nabla^2\phi, \tag{8.33}$$

básicamente por la misma definición de los operadores. El caso del rotacional de un gradiente también es bastante directo:[2]

$$(\vec{\nabla}\times\vec{\nabla}\phi)^\mu \;=\; \varepsilon^{\mu\nu\rho}\,\nabla_\nu\nabla_\rho\phi \;=\; \frac{1}{2}\varepsilon^{\mu\nu\rho}\,[\nabla_\nu,\nabla_\rho]\phi \;=\; 0, \tag{8.34}$$

El caso de la divergencia de un rotacional es un poco más sútil. El conocido resultado de $\mathbb{R}^N$ generaliza a variedades curvas, pero gracias a la identidad de Bianchi (8.9):

$$\vec{\nabla}\cdot(\vec{\nabla}\times\vec{A}) \;=\; \varepsilon^{\mu\nu\rho}\,\nabla_\mu\nabla_\nu A_\rho \;=\; \frac{1}{2}\varepsilon^{\mu\nu\rho}\,[\nabla_\mu,\nabla_\nu]A_\rho \;=\; \frac{1}{2}\varepsilon^{\mu\nu\rho}\,R_{\mu\nu\rho}{}^\lambda A_\lambda \;=\; 0. \tag{8.35}$$

Finalmente el rotacional de un rotacional sí da un resultado distinto cuando hay curvatura:

$$\begin{aligned}(\vec{\nabla}\times\vec{\nabla}\times\vec{A})^\mu \;&=\; \varepsilon^{\mu\nu\rho}\varepsilon_{\rho\lambda\sigma}\nabla_\nu\nabla^\lambda A^\sigma\\ &=\; \nabla_\nu\nabla^\mu A^\nu \;-\; \nabla_\nu\nabla^\nu A^\mu\\ &=\; g^{\mu\lambda}\Big[\nabla_\lambda\nabla_\nu A^\nu \;+\; R_{\nu\lambda\sigma}{}^\nu A^\sigma\Big] \;-\; \nabla^2 A^\mu\\ &=\; \nabla^\mu\nabla_\nu A^\nu \;-\; \nabla^2 A^\mu \;-\; R^{\mu\sigma}A_\sigma,\end{aligned} \tag{8.36}$$

donde en la segunda igualdad hemos usado la identidad para la contracción de dos tensores de Levi-Civita,

$$\varepsilon^{\mu\nu\rho}\varepsilon_{\lambda\sigma\rho} \;=\; \delta_\lambda^\mu\delta_\sigma^\nu \;-\; \delta_\lambda^\nu\delta_\sigma^\mu. \tag{8.37}$$

[2]Por lo menos en el caso de que la conexión sea Levi-Civita, al que nos limitaremos aquí. Dejamos como ejercicio derivar el equivalente para conexiones arbitrarias.

El rotacional de un rotacional tiene por lo tanto un término extra con respecto al resultado en $\mathbb{R}^3$, proporcional al tensor de Ricci, a causa del intercambio de dos derivadas covariantes en la tercera igualdad, para poder escribir el primer término como el gradiente de una divergencia.

A estas alturas es natural preguntarnos por qué la identidad (8.36) depende de la geometría del espacio considerado, pero (8.34) y (8.35) son independientes de la métrica. La respuesta es relativamente sencilla si recordamos los teoremas de Stokes (1.25): la integral de un rotacional en una superficie S se puede escribir como la integral de linea sobre la curva $\gamma = \partial S$ que bordea la superficie, de modo que para nuestro caso tenemos que

$$\int_S (\vec{\nabla} \times \vec{\nabla}\phi) \cdot \vec{n}\, d^2x \;=\; \oint_{\partial S} \vec{\nabla}\phi \cdot d\vec{r}. \tag{8.38}$$

Sin embargo esta última expresion es cero, por lo menos en espacios simplemente conexos (espacios donde todos las curvas con contraíbles a un punto): la integral de linea de un gradiente de un campo es igual a la diferencia de los valores el campo en los extremos de la curva. Pero al ser $\gamma = \partial S$ una curva cerrada, la integral (8.38) vale cero. Dicho de otra manera: en un espacio simplemente conexo, la curva $\gamma = \partial S$ se puede contraer a un solo punto, por lo que la integral no puede tener un valor finito.

Del mismo modo el teorema de Stokes convierte la integral de volumen V de una divergencia en la integral en la superficie $S = \partial V$ que rodea el volumen V. Si entonces la divergencia actúa sobre un rotacional, tenemos que

$$\int_V \vec{\nabla} \cdot (\vec{\nabla} \times \vec{A})\, d^3x \;=\; \oint_{\partial V} (\vec{\nabla} \times \vec{A}) \cdot d^2x, \tag{8.39}$$

lo que a su vez por el teorema de Stokes sería la integral de $\vec{A}$ sobre la curva que bordea la superficie $S = \partial V$. Sin embargo, al ser ∂V una superficie cerrada, no tiene una curva $\gamma = \partial S = \partial\partial V$ que la rodea y la integral (8.39) es trivial.

En otras palabras, la identidades (8.34) y (8.35) son propiedades topológicas de espacios simplemente conexos y por lo tanto no dependen de la geometría (métrica) del espacio. Son manifestaciones del conocido teorema topológico que "la frontera de una frontera es cero". Sin embargo, para construir el operador $\vec{\nabla} \times \vec{\nabla} \times \vec{A}$ es preciso contraer dos tensores de Levi-Civita a través de la métrica, de modo que aquí sí aparecen la propiedades geométricos del espacio.

8.5. Las coordenadas localmente inerciales

En cada punto p de una variedad diferenciable hay un sistema local de coordenadas que juega un papel fundamental para la relatividad general: las *coordenadas localmente inerciales*. Estas coordenadas están caracterizadas por el hecho de que en el punto p la métrica toma la forma de una métrica plana y las primeras derivadas se anulan,[3]

$$g_{ij}(p) \stackrel{*}{=} \delta_{ij}, \qquad\qquad \partial_k g_{ij}(p) \stackrel{*}{=} 0, \tag{8.40}$$

donde hemos introducido el símbolo "$\stackrel{*}{=}$" para indicar que una igualdad sólo es válida en un sistema de coordenadas específico, no en general.

Para demostrar esta propiedad, basta con contar el número de componentes independientes de la métrica y compararlo con el número de grados de libertad que tenemos para elegir las coordenadas. Si $g_{\mu\nu}(x)$ es la métrica una variedad arbitraria en coordenadas x^μ, entonces en un

[3]Trataremos aquí el caso riemanniano, donde la métrica plana en coordenadas cartesianas toma la forma $\delta_{\mu\nu}$. El caso lorentziano, donde la métrica coge la forma $\eta_{\mu\nu}$ es completamente análogo y no merece la pena hacer explicitamente.

pequeño entorno alrededor de un punto p podemos expandirla como

$$g_{\mu\nu}(x) \ \approx\ g_{\mu\nu}(p) \ +\ \partial_\rho g_{\mu\nu}(p)\, x^\rho \ +\ \frac{1}{2}\partial_\rho\partial_\lambda g_{\mu\nu}(p)\, x^\rho\, x^\lambda \ +\ \ldots \tag{8.41}$$

donde por simplicidad (pero sin pérdida de generalidad) hemos elegido $x^\mu(p) = 0$. Del mismo modo en las coordenadas localmente inerciales y^i tenemos que

$$g_{ij}(y) \ \approx\ g_{ij}(p) \ +\ \partial_k g_{ij}(p)\, y^k \ +\ \frac{1}{2}\partial_k\partial_l g_{ij}(p)\, y^k\, y^l \ +\ \ldots \tag{8.42}$$

Las dos expresiones de la métrica obviamente están relacionadas a través de un cambio de coordenadas (6.29),

$$g_{ij}(y) \ =\ \frac{\partial x^\mu}{\partial y^i}\,\frac{\partial x^\nu}{\partial y^j}\, g_{\mu\nu}(x), \tag{8.43}$$

donde a su vez podemos expandir el cambio de coordenadas $x^\mu = x^\mu(y)$ en una serie de Taylor como

$$x^\mu \ \approx\ \frac{\partial x^\mu}{\partial y^i}(p)\, y^i \ +\ \frac{\partial^2 x^\mu}{\partial y^i \partial y^j}(p)\, y^i\, y^j \ +\ \frac{\partial^3 x^\mu}{\partial y^i \partial y^j \partial y^k}(p)\, y^i\, y^j\, y^k \ +\ \ldots \tag{8.44}$$

Sustituyendo (8.41) , (8.42) y (8.44) en (8.43) obtenemos (esquemáticamente) que (ejerc.)

$$\begin{aligned} g'(p) \ &+\ \partial g'(p)\, y \ +\ \partial\partial g'(p)\, y^2 \ \approx\ \Big[\frac{\partial x}{\partial y}\frac{\partial x}{\partial y}\, g\Big](p) \ +\ \Big[\frac{\partial x}{\partial y}\frac{\partial x}{\partial y}\frac{\partial x}{\partial y}\,\partial g \ +\ \frac{\partial x}{\partial y}\frac{\partial^2 x}{\partial y\partial y}\, g\Big](p)\, y \qquad (8.45)\\ &+\ \Big[\frac{\partial x}{\partial y}\frac{\partial x}{\partial y}\frac{\partial x}{\partial y}\frac{\partial x}{\partial y}\,\partial\partial g \ +\ \frac{\partial x}{\partial y}\frac{\partial x}{\partial y}\frac{\partial^2 x}{\partial y\partial y}\,\partial g \ +\ \Big(\frac{\partial x}{\partial y}\frac{\partial^3 x}{\partial y\partial y\partial y} + \frac{\partial^2 x}{\partial y\partial y}\frac{\partial^2 x}{\partial y\partial y}\Big) g\Big](p)\, y\, y \ +\ \ldots \end{aligned}$$

donde el acento sólo indica que escribimos la métrica en coordenadas y^i. Comparando el lado izquierdo con el lado derecho de esta ecuación tenemos entonces que

$$\begin{aligned} g'(p) \ &=\ \Big[\frac{\partial x}{\partial y}\frac{\partial x}{\partial y}\, g\Big](p), \qquad (8.46)\\ \partial g'(p) \ &=\ \Big[\frac{\partial x}{\partial y}\frac{\partial x}{\partial y}\frac{\partial x}{\partial y}\,\partial g \ +\ \frac{\partial x}{\partial y}\frac{\partial^2 x}{\partial y\partial y}\, g\Big](p),\\ \partial\partial g'(p) \ &=\ \Big[\frac{\partial x}{\partial y}\frac{\partial x}{\partial y}\frac{\partial x}{\partial y}\frac{\partial x}{\partial y}\,\partial\partial g \ +\ \frac{\partial x}{\partial y}\frac{\partial x}{\partial y}\frac{\partial^2 x}{\partial y\partial y}\,\partial g \ +\ \Big(\frac{\partial x}{\partial y}\frac{\partial^3 x}{\partial y\partial y\partial y} + \frac{\partial^2 x}{\partial y\partial y}\frac{\partial^2 x}{\partial y\partial y}\Big) g\Big](p). \end{aligned}$$

El resto es combinatoria: $g'_{ij}(p)$ es un tensor simétrico, de modo que en una variedad N-dimensional tiene $\frac{1}{2}N(N+1)$ componentes independientes. Por otro lado, la matriz $\partial x^\mu/\partial y^i$ tiene N^2 entradas independientes, asi que, sea lo que sea el valor de $g_{\mu\nu}(p)$, siempre podemos elegir los $\partial x^\mu/\partial y^i$ tal que $g_{ij}(p) = \delta_{ij}$.[4]

Una vez elegidas las $\partial x^\mu/\partial y^i$, las primeras derivadas de la métrica $\partial_k g_{ij}(p)$ están determinadas por las segundas derivadas del cambio de coordenadas $\partial^2 x^\mu/\partial y^i \partial y^j$. No es difícil ver que hay $\frac{1}{2}N^2(N+1)$ componentes $\partial_k g_{ij}(p)$ y exactamente la misma cantidad de entradas $\partial^2 x^\mu/\partial y^i \partial y^j$, de modo que siempre hay suficiente libertad para elegir los $x^\mu(y^i)$ tal que $\partial_k g_{ij}(p) = 0$.

Sin embargo, los números ya no cuadran cuando miramos las segundas derivadas de la métrica: dado que $\partial_k\partial_l g_{ij}$ es simétrico en tanto en i y j, como en k y l, tenemos $\frac{1}{4}N^2(N+1)^2$ segundas derivadas independientes de g_{ij}, mientras sólo hay $\frac{1}{6}N^2(N+1)(N+2)$ componentes $\partial^3 x^\mu/\partial y^i \partial y^j \partial y^k$. Podemos por lo tanto igualar a cero una parte de las entradas de $\partial_k\partial_l g_{ij}$, pero siempre quedarán $\frac{1}{12}N^2(N^2-1)$ componentes que no podemos elegir. Veremos en breve que no

[4] Se puede demostrar que los restante $\frac{1}{2}N(N-1)$ componentes de $\partial x^\mu/\partial y^i$ son exactamente las componentes de las transformaciones ortogonales $SO(N)$ que preservan la forma de la métrica δ_{ij} ($\eta_{\mu\nu}$).

es casualidad que este número es precisamente el número de componentes independientes del tensor de Riemann.

Resumiendo, para cualquier punto de una variedad siempre podemos elegir unas coordenadas tal que la métrica en una pequeña región alrededor de este punto sea de la forma

$$g_{ij}(x) \stackrel{*}{=} \delta_{ij} + \mathcal{O}(\partial^2 g\, y^2). \tag{8.47}$$

En el fondo no es demasiado profundo lo que acabamos de hacer. Sabemos que una variedad $\mathcal{M}^N$ es un espacio que localmente parece $\mathbb{R}^N$, porque en cada punto p podemos definir el espacio tangente $T_p(\mathcal{M})$, que es isomorfo a $\mathbb{R}^N$. Lo que quiere decir la expresión (8.47) es que en una región suficientemente pequeña alrededor de p, el espacio tangente $T_p(\mathcal{M})$ es una buena aproximación de la variedad y las coordenadas localmente inerciales son las coordenadas cartesianas en el espacio tangente. Además vemos que la diferencia local entre la metrica g_{ij} de la variedad y la métrica δ_{ij} ($\eta_{\mu\nu}$) del plano tangente es un efecto de segunda orden. Matemáticamente hablando g_{ij} será de la forma δ_{ij} ($\eta_{\mu\nu}$) solamente en el punto p, ni siquiera en una pequeña región, ya que las derivadas segundas no son cero. Pero desde el punto de vista físico, siempre habrá una pequeña región donde este efecto de segundo orden es demasiado pequeño para ser detectado y en la cual no se nota la diferencia entre la variedad curva y el espacio tangente. La extensión de esta región dependerá de la sensibilidad de los aparatos experimentales usados.

Nótese que las coordenadas localmente inerciales serán diferentes en puntos diferentes, ya que el cambio de coordenadas para ir a estas coordenadas varía de punto en punto. Por lo tanto siempre es posible llevar la métrica a la forma (8.47) en cualquier punto p, pero en general no es todos los puntos a la vez (sino el espacio sería plano). Esto coincide con la idea rudimentaria que teniamos de una variedad: un conjunto de parches que son localmente planos y "pegados" de forma suave tal que forma un espacio que globalmente no es plano.

El hecho de que en coordenadas localmente inerciales las primeras derivadas de la métrica son idénticamente cero en un punto p, implica que en estas coordenadas los símbolos de Christoffel $\Gamma^\rho_{\mu\nu}$ son cero en este punto, ya que éstas se construye directamente de las derivadas de la métrica.

El cambio de coordenadas es en realidad el mismo que el que hemos construido antes, pero es instructivo ver cómo funciona para los símbolos de Christoffel a través de las reglas de transformación (7.15). Considera el cambio de coordenadas

$$y^i = x^\mu\, \delta^i_\mu + \frac{1}{2} M^i_{\mu\nu} x^\mu x^\nu, \tag{8.48}$$

donde $M^i_{\mu\nu}$ son N^3 constantes que quedan por determinar. Tomemos, sin pérdida de generalidad, como el origen $x^\mu = 0$ del sistema de referencia el punto p donde queremos construir el sistema localmente inercial. Bajo este cambio de coordenadas la conexión transforma en el punto $x^\mu = 0$ como

$$\Gamma^k_{ij}(p) = \delta^\mu_i\, \delta^\nu_j\, \delta^k_\rho\, \Gamma^\rho_{\mu\nu}(p) - M^k_{\mu\nu}\, \delta^\mu_i\, \delta^\nu_j. \tag{8.49}$$

Con la elección apropiada de las constantes $M^i_{\mu\nu} = \Gamma^\rho_{\mu\nu}(p)\, \delta^i_\rho$, podemos conseguir que en el punto p la conexión en las coordenadas y^i tengan el valor $\Gamma^k_{ij}(p) \stackrel{*}{=} 0$. Nótese otra vez que el cambio de coordenadas con esta elección de las $M^i_{\mu\nu}$ sólo anula la conexión en el punto p. Para poder anular la conexión en un punto $q \neq p$ hará falta otra elección de las constantes, es decir, $M^i_{\mu\nu} = \Gamma^\rho_{\mu\nu}(q)\, \delta^i_\rho$.

Nótese que sólo hemos conseguido imponer una condición sobre el valor de los símbolos de Christoffel en un punto p, pero no sobre sus derivadas. Por lo tanto, aunque la métrica tome la forma (8.47) y aunque $\Gamma^k_{ij}(p) \stackrel{*}{=} 0$, el tensor de Riemann sigue teniendo una forma no-trivial

$$R_{ijk}{}^l(p) \stackrel{*}{=} \partial_i \Gamma^l_{jk}(p) - \partial_j \Gamma^l_{ik}(p). \tag{8.50}$$

Son precisamente las segundas derivadas de g_{ij}, que no logramos anular en el cambio a coordenadas localmente inerciales que contribuyen al tensor de Riemann y las derivadas de los símbolos

de Christoffel. Nótese que el número de componentes independientes del tensor de Riemann es igual al número de segundas derivadas $\partial_k \partial_l g_{ij}$ que no podemos anular.

Finalmente, nótese qué esta propiedad no funcionaría en presencia de torsión: el cambio de coordenadas (8.48) es simétrico en los índices μ y ν, de modo que en (8.49) el cambio de coordenadas sólo podrá deshacer la parte simétrica de $\Gamma_{\mu\nu}^{\rho}$. Pero también hay otra manera de verlo: si la conexión no fuera simétrica, la torsión $T_{\mu\nu}^{\rho} = \Gamma_{\mu\nu}^{\rho} - \Gamma_{\nu\mu}^{\rho}$ sería diferente de cero en un sistema de coordenadas arbitrario. Pero el caracter tensorial de la torsión impide que en este caso exista un sistema de coordenadas específico donde la torsión (y por lo tanto la conexión) sea cero, aunque sea en un punto. La única posibilidad para poder igualar a cero las conexiónes es que la torsión sea cero, es decir, cuando la conexión sea simétrica.

Las coordenadas localmente inerciales juegan un papel importante en la relatividad general. En la sección 5.4 dijimos que, gracias el Principio de Equivalencia, un observador en caída libre en un campo gravitatorio se puede creer un observador inercial en ausencia de gravedad, por lo menos en una región pequeña. Efectivamente, en este caso el observador no es capaz de ver la diferencia entre la variedad curva y el espacio tangente en el punto donde está. Además las coordenadas localmente inerciales son localmente idénticas a las coordenadas cartesianas, de modo que se creerá un observador inercial en el espacio de Minkowski.

8.6. Geodésicas con Levi-Civita

Otro sitio donde se nota la ventaja de usar la conexión de Levi-Civita es en el estudio de las geodésicas. En el Capítulo 7 hemos visto que en una variedad curva se pueden definir dos tipos de geodésicas: las afines y las métricas. Las geodésicas afines, definidas por la ecuación (7.42)

$$\ddot{x}^{\rho} \;+\; \Gamma_{\mu\nu}^{\rho}\, \dot{x}^{\mu}\, \dot{x}^{\nu} \;=\; 0, \tag{8.51}$$

eran aquellas curvas donde el vector tangente se mantenía paralelo a si mismo bajo transporte paralelo a lo largo de la curva. Por otro lado, las geodésicas métricas, las curvas más cortas entre dos puntos, estaban definidas por

$$\ddot{x}^{\rho} \;+\; \left\{{}^{\;\rho}_{\mu\nu}\right\} \dot{x}^{\mu}\, \dot{x}^{\nu} \;=\; 0. \tag{8.52}$$

Ahora, la expresión (8.7) para la conexión de Levi-Civita en función de la métrica coincide exactamente con la expresión de los símbolos de Christoffel (7.49) que aparece en la fórmula de las geodésicas métricas,

$$\Gamma_{\mu\nu}^{\rho} \;=\; \frac{1}{2}\, g^{\rho\lambda} \Big(\partial_{\mu} g_{\lambda\nu} \;+\; \partial_{\nu} g_{\mu\lambda} \;-\; \partial_{\lambda} g_{\mu\nu} \Big) = \left\{{}^{\;\rho}_{\mu\nu}\right\} \tag{8.53}$$

de modo que en variedades equipadas con la conexión de Levi-Civita, las dos ecuaciones (8.51) y (8.52) coinciden y ya no hay diferencia entre las geodésicas afines y las geodésicas métricas.

Desde el punto de vista físico, esta simplificación tiene mucho sentido. Dado que el vector tangente a una curva $\dot{x}^{\mu}$ es la velocidad de la partícula que sigue dicha curva, las geodésicas afines son trayectorias de partículas no-aceleradas, ya que su la velocidad es constante a lo largo de la curva: $\dot{x}^{\mu} \nabla_{\mu} \dot{x}^{\rho} = 0$. Por otro lado, las geodésicas métricas, al venir de un principio variacional, están íntimamente ligadas con las ecuaciones de movimiento en el formalismo lagrangiano. Efectivamente, la ecuación de la geodésica (8.52) se puede obtener equivalentemente[5] tras aplicar la ecuación de Euler-Lagrange a la funcional

$$S \;=\; \int d\tau\; g_{\mu\nu}(x)\, \dot{x}^{\mu}(\tau)\, \dot{x}^{\nu}(\tau). \tag{8.54}$$

[5]En la práctica, al multiplicar (7.46) por $2L$ para obtener la ecuación (7.47), estamos haciendo la variación de la funcional (8.54).

Esta es, módulo un factor $\frac{1}{2}m_0$, la energía cinética de una partícula en un espacio descrita por una métrica $g_{\mu\nu}$ y las geodésicas métricas corresponden por lo tanto a las trayectorias de partículas libres (es decir: en ausencia de un potencial).

Ahora, en el capítulo anterior hemos visto que en una variedad equipada con una conexión arbitraria las geodésicas métricas no necesariamente coinciden con las geodésicas afines y uno se podría preguntar qué trayectoria seguiría una partícula libre en este caso: ¿una curva no-acelerada, o una trayectoria que minimiza la acción? La conexión de Levi-Civita resuelva esta paradoja de manera elegante, al identificar los dos tipos de curvas.

La relación entre la conexión de Levi-Civita y el formalismo lagrangiano va más lejos que solamente el caso de las partículas libres. En general, para un lagrangiano $\mathcal{L} = T - V$ con energía cinética $T = T(q, \dot{q})$ y potencial $V = V(q)$, la ecuación de Euler-Lagrange (7.46) se puede escribir como

$$\frac{d}{d\tau}\Big(\frac{\partial T}{\partial \dot{q}^a}\Big) \;-\; \frac{\partial T}{\partial q^a} \;=\; -\,\frac{\partial V}{\partial q^a}. \tag{8.55}$$

Nótese que el primer término a la izquierda da lugar a un término de la aceleración, $m_0\ddot{q}^a$, y que podemos identificar el térmimo a la derecha como la fuerzas físicas en coordenadas generalizadas, $F^a = -\partial V/\partial q^a$, cosa que uno esperaría en una Segunda Ley de Newton. Pero lo interesante en este contexto es que el primer y el segundo término a la izquierda dan lugar a términos proporcionales a $\dot{q}^a\dot{q}^b$, donde los constantes de proporcionalidad son precisamente los símbolos de Christoffel (8.53). La ecuación entonces toma la forma

$$m_0\ddot{q}^a \;+\; \Gamma^a_{bc}\,\dot{q}^b\,\dot{q}^c \;=\; -\partial^a V(q). \tag{8.56}$$

En el formalismo lagrangiano, las ligaduras al que está sometido el sistema definen el llamado *espacio de configuraciones*, en que el se mueve el sistema. Este espacio de configuraciones es una variedad (en general curva), parametrizada por las coordenadas generalizadas q^a. Vemos por lo tanto que la dinámica del sistema, descrita por la ecuación de Euler-Lagrange, no es más que la Segunda Ley de Newton, escrita en términos de las coordenadas curvilíneas q^a. En ausencia de fuerzas físicas, el sistema seguirá una geodésica dentro del espacio de configuraciones, pero para potenciales no-triviales, la trayectoria se desviará de ésta por el término inhomogeneo en (8.56). En el Cápítulo 10 veremos que las ecuaciones de movimiento de una partícula de prueba en la relatividad general son de este tipo.

8.7. Desviación geodésica

En la geometría diferencial hay una relación importante que describe la evolución de familias de geodésicas: en una variedad curva, las geodésicas que inicialmente tienen direcciones paralelas, en general no seguirán paralelas, sino se alejarán o se acercarán entre ellos. La ecuación de la *desviación geodésica* describe este efecto en términos de la curvatura de la variedad.

Considera una familia de geodésicas $\gamma_\sigma(\tau)$, donde el parametro σ identifica las geodésicas específicas dentro de la familia y τ parametriza los puntos a lo largo de cada geodésica (véase Figura 8.1). La familia de geodésicas define una superficie bidimensional $x^\mu(\sigma, \tau)$ en la variedad, en la que podemos usar σ y τ como coordenadas naturales.

Definimos ahora las siguientes dos vectores tangentes a la superficie:

$$u^\mu \;=\; \frac{\partial x^\mu}{\partial \tau}, \qquad\qquad s^\mu \;=\; \frac{\partial x^\mu}{\partial \sigma}. \tag{8.57}$$

Definidos de esta manera, los vectores u^μ y s^μ tienen una interpretación muy natural: u^μ es la velocidad a lo largo de la geodésica y s^μ mide la separación entre dos geodésicas cercanas. El hecho

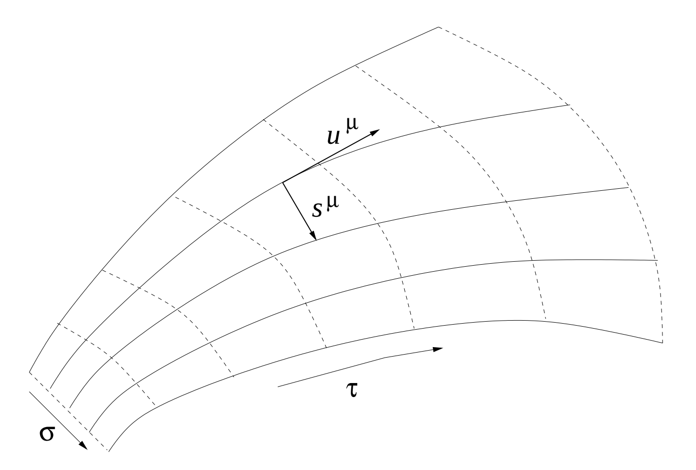


Figura 8.1: *Una familia de geodésicas $\gamma_\sigma(\tau)$, donde τ es el parámetro afín a lo largo de cada geodésica y σ el parámetro que identifica cada geodésica dentro de la familia. La familia de geodésicas define una superficie bidimensional dentro de la variedad y en cada punto se pueden definir los vectores u^μ y s^μ que representan respectivamente la velocidad a lo largo de la geodésica y la separación entre geodésicas cercanas.*

de que u^μ y s^μ formen una base en la superficie, implica que satisfacen la siguiente condición:[6]

$$u^\rho \nabla_\rho s^\mu \;=\; s^\rho \nabla_\rho u^\mu. \tag{8.58}$$

Aparte de la velocidad u^μ a lo largo de la geodésicas y la separación s^μ entre geodésicas cercanas, podemos definir los vectores V^μ y A^μ como

$$V^\mu \;=\; u^\rho \nabla_\rho s^\mu, \qquad\qquad A^\mu \;=\; u^\rho \nabla_\rho V^\mu, \tag{8.59}$$

que representan cuánto cambia la separación entre distintas geodésicas a lo largo de la curva a primer y segundo orden. En cierto modo V^μ es la velocidad de receso, que dice cuánto aumenta la distancia entre las geodésicas según vayan avanzando, y A^μ la aceleración relativa entre geodésicas, que mide los cambios en la velocidad de recesión. Ojo, no se deben confundir ninguna de estos conceptos con la (cuadri)velocidad y la aceleración de las partículas que se mueven a lo largo de las geodésicas.

Ahora, la desviación geodésica es una relación entre la aceleración relativa y la curvatura de la variedad. En particular para A^μ tenemos que

$$\begin{aligned}
A^\lambda &= u^\nu \nabla_\nu (u^\mu \nabla_\mu s^\lambda) \\
&= u^\nu \nabla_\nu (s^\mu \nabla_\mu u^\lambda) \\
&= (u^\nu \nabla_\nu s^\mu) \nabla_\mu u^\lambda \;+\; u^\nu s^\mu \nabla_\nu \nabla_\mu u^\lambda \\
&= (s^\nu \nabla_\nu u^\mu) \nabla_\mu u^\lambda \;+\; u^\nu s^\mu \nabla_\mu \nabla_\nu u^\lambda \;+\; u^\nu s^\mu R_{\nu\mu\rho}{}^\lambda u^\rho \\
&= (s^\nu \nabla_\nu u^\mu) \nabla_\mu u^\lambda \;+\; s^\mu \nabla_\mu (u^\nu \nabla_\nu u^\lambda) \;-\; (s^\mu \nabla_\mu u^\nu) \nabla_\nu u^\lambda \;-\; R_{\mu\nu\rho}{}^\lambda s^\mu u^\nu u^\rho \\
&= -R_{\mu\nu\rho}{}^\lambda s^\mu u^\nu u^\rho
\end{aligned} \tag{8.60}$$

donde en las segunda y la cuarta igualdad hemos usado la propiedad (8.58), en la tercera y la quinta la regla de Leibniz para derivadas covariantes y en la sexta que $\gamma_\sigma(\tau)$ es una geodésica y que por lo tanto $u^\nu \nabla_\nu u^\lambda = 0$.

[6]En el Capítulo **??** veremos que esto quiere decir que el conmutador entre los dos vectores es cero: $[u, s]^\mu = 0$.

Aquí hemos derivado el resultado para una familia de geodésicas, parametrizado por un solo parámetro, pero se puede considerar también en caso de una congruencia de geodéscias. Si $\dot{x}^\mu$ es el vector tangente a las geodésicas y δx^μ el vector de desplazamiento, en general la fórmula de la desviación geodésica viene dada por

$$\frac{\nabla^2 \delta x^\lambda}{d\tau^2} \;=\; -R_{\mu\nu\rho}{}^\lambda\, \delta x^\mu\, \dot{x}^\nu\, \dot{x}^\rho. \tag{8.61}$$

Vemos por lo tanto que las geodésicas que inicialmente tienen direcciones paralelas, experimentan aceleraciones relativas debido a la curvatura del espacio, que hacen que las geodésicas convergen o divergen. Sólo en el espacio plano, $R_{\mu\nu\rho}{}^\lambda = 0$, la desviación geodésica es cero y las geodésicas se mantienen paralelas. En cierto modo, esto es la formulación en lenguaje de geometría diferencial de la versión de John Playfair del Quinto Postulado de Euclides, que hemos visto en la sección 6.1.

Desde el punto de vista físico, la desviación geodésica describe los efectos de marea. En el Capítulo 10 veremos que las ecuaciones de Einstein implican que el espaciotiempo se curva en presencia de materia y la desviación geodésica dice que dos geodésicas notarán una aceleración relativa entre ellos debido a la curvatura del espaciotiempo. Las distintas partículas de una configuración de partículas de prueba seguirán trayectorias distintas y la configuración se deformará, lo que un observador externo interpretará como fuerzas de marea. Los efectos de marea son más grandes, cuanto más grande la curvatura, por eso típicamente aparecen donde el gradiente del campo gravitatorio en muy grande, como cerca del centro de agujeros negros.

8.8. Ejemplo concreto: geodésicas en la $\mathbb{S}^2$

Ilustremos estos conceptos con el ejemplo concreto de la esfera bidimensional $\mathbb{S}^2$ en coordenadas azimutales,

$$ds^2 \;=\; R_0^2 \Big(d\theta^2 \;+\; \sin^2\theta d\varphi^2\Big). \tag{8.62}$$

En (8.15) hemos visto que los símbolos de Christoffel no-triviales son

$$\Gamma_{\theta\varphi}^{\varphi} = \Gamma_{\varphi\theta}^{\varphi} = \operatorname{cotg}\theta, \qquad\qquad \Gamma_{\varphi\varphi}^{\theta} = -\sin\theta\cos\theta, \tag{8.63}$$

de modo que las distintas componentes de la ecuación de la geodésica (8.51) vienen dadas por

$$\ddot{\theta} \;-\; \sin\theta\,\cos\theta\,\dot{\varphi}^2 \;=\; 0, \qquad\qquad \ddot{\varphi} \;+\; 2\operatorname{cotg}\theta\,\dot{\theta}\,\dot{\varphi} \;=\; 0. \tag{8.64}$$

Además podemos añadir la condición (7.53)

$$R_0^2\,\dot{\theta}^2 \;+\; R_0^2\,\sin^2\theta\,\dot{\varphi}^2 \;=\; 1, \tag{8.65}$$

para geodésicas espaciales. En realidad, en una variedad riemanniana, donde todas las geodésicas con espaciales, esta condición es superflua, pero nos dice que la parametrización de la geodésica es tal que la norma de la velocidad a lo largo de la curva siempre es la unidad. Veremos que parametrizar la geodésica de esta manera nos ayudará a resolver el sistema de ecuaciones (8.64).

Una primera observación es que se puede integrar la segunda ecuación de (8.64) a

$$\dot{\varphi} \;=\; \ell\sin^{-2}\theta, \tag{8.66}$$

donde ℓ es una constante de integración, que representa el momento angular. Sustituyendo esto en la condición (8.65) tenemos que

$$\dot{\theta}^2 \;=\; R_0^{-2}\Big[1 \;-\; \ell^2\,R_0^2\,\sin^{-2}\theta\Big]. \tag{8.67}$$

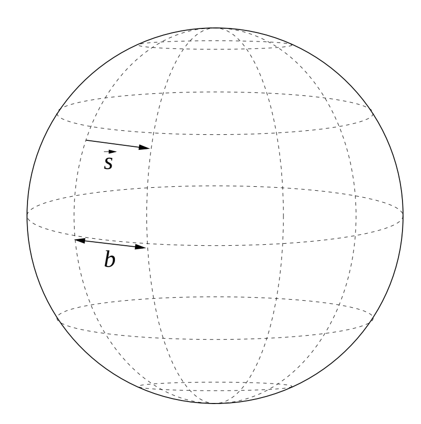


Figura 8.2: *La desviación geodésica en la esfera bidimensional: la separación $\vec{s}$ entre dos geodéscias varía entre un máximo y cero, ya que cualesquieras dos círculos máximos se cruzan en dos exactamente puntos.*

Las dos ecuaciones para $\theta = \theta(\tau)$ y $\varphi = \varphi(\tau)$ se pueden combinar en una sola ecuación para $\theta = \theta(\varphi)$, (ejerc.)

$$\frac{d\theta}{d\varphi} = R_0^{-1}\ell^{-1}\sqrt{\sin^4\theta - R_0^2\,\ell^2\,\sin^2\theta} \tag{8.68}$$

que tiene como solución (ejerc.)

$$\sin(\varphi - \varphi_0) = \frac{R_0\,\ell}{\sqrt{1 - R_0^2\ell^2}}\,\mathrm{cotg}\,\theta. \tag{8.69}$$

Para dar una interpretación geométrica de lo que representa esta ecuación, es útil escribirla como

$$-\sin\varphi_0\sin\theta\cos\varphi + \cos\varphi_0\sin\theta\sin\varphi - \frac{R_0\,\ell}{\sqrt{1 - R_0^2\ell^2}}\cos\theta = 0, \tag{8.70}$$

que, multiplicado por r e identificando $A = -\sin\varphi_0$, $B = \cos\varphi_0$ y $C = R_0\ell(1 - R_0^2\ell^2)^{-\frac{1}{2}}$, no es más que la ecuación en coordenadas esfércias de un plano que pasa por el origen de $\mathbb{R}^3$:

$$Ax + By + Cz = 0. \tag{8.71}$$

En otras palabras la solución (8.69) de la geodésica en $\mathbb{S}^2$ es la intersección de esa esfera con un plano que pasa por el origen, lo que viene a ser un círculo máximo (un círculo que tiene como centro el centro de la esfera), donde los constantes de integración ℓ y φ_0 determinan la inclinación y la orientación de ese círculo máximo con respecto al ecuador.

Una vez determinadas las geodésicas en la esfera bidimensional, es útil calcular la desviación geodésica. Tomamos por simplicidad como geodésica de referencia uno de los meridianos, es decir un círculo máximo que pasa por los polos, y como segunda geodéscia otro meridiano que en el ecuador está separado por una distancia b del primero (véase Figura 8.2).

La ecuación de la desviación geodésica (8.61) impone entonces las siguientes condiciones sobre las componentes del vector de separación $\vec{s}$:

$$\frac{\nabla^2 s^\theta}{d\tau^2} = -R_{\varphi\theta\theta}{}^{\theta}\,s^\varphi\,u^\theta\,u^\theta, \qquad \frac{\nabla^2 s^\varphi}{d\tau^2} = -R_{\varphi\theta\theta}{}^{\varphi}\,s^\varphi\,u^\theta\,u^\theta, \tag{8.72}$$

Parametrizando las geodésicas como $\gamma_s(\tau) = \tau = R_0^{-1}\theta$, tenemos entonces que $\vec{u} = R_0^{-1}\vec{e}_\theta$, de modo que con las expresiones (8.16) para el tensor de Riemann de la $\mathbb{S}^2$, estas ecuaciones se reducen al siguiente sistema de ecuaciones diferenciales desacopladas:

$$\frac{d^2 s^\theta}{d\theta^2} = 0, \qquad \frac{d^2 s^\varphi}{d\theta^2} = -s^\varphi. \tag{8.73}$$

Se ve fácilmente que la solución que corresponde a las condiciones iniciales

$$s^\theta(\theta = \pi/2) = 0, \qquad\qquad s^\varphi(\theta = \pi/2) = b, \tag{8.74}$$

viene dada por

$$\vec{s}(\theta) \;=\; b\,\sin\theta\,\vec{e}_\varphi. \tag{8.75}$$

En otras palabras, las geodésicas tienen una separación máxima b en el ecuador, empiezan a convergir según se acercan a los polos y intersectan en los polos. Gracias a la homogeneidad de la esfera, este comportamiento no sólo es valido para meredianos, sino para todas las geodéscias. Efectivamente, sabemos que en una esfera cualesquiera dos círculos máximos se intersectan en exactamente dos puntos antipodales.

La desviación geodésica juega un papel importante en la transición conceptual entre la gravedad newtoniana y la relatividad general, ya que explica la "misteriosa fuerza a distancia" en términos de la curvatura del espaciotiempo. Discutiremos la relación entre ambas cosas en mucho detalle en la Parte III, pero aquí podemos dar un ejemplo intuitivo. Imaginemos dos aviones que salen simultaneamente desde dos puntos distintos en el ecuador, con rumbo al Polo Norte. Si los dos siguen geodésicas, se irán acercando poco a poco y finalmente chocarán en el polo. Desde el punto de vista de geometría diferencial este análisis es obvio, pero la percepción que tienen los dos pilotos es distinta. Desde sus aviones, éstos no se dan cuenta de la curvatura de la Tierra y creen que siguen lineas rectas en un espacio plano, mientras que en realidad van "saltando de espacio tangente en espacio tangente" y que están adaptando constantemente sus coordenadas localmente inerciales, creyendo que es un único sistema de coordenadas cartesianas. Por lo tanto, para explicar dentro de estas coordenadas cartesianas la aceleración relativa entre los dos aviones, tienen que invocar una misteriosa fuerza que entre los aviones, que hace que se atraen y que salen de su rutas inicialmente paralelas.

La idea de la interacción gravitatoria entre partículas distantes es en realidad una manifestación de la curvatura del espaciotiempo es la esencia de la relatividad general. En los siguientes capítulos investigaremos la relación entre ambos conceptos en detalle.

Parte III

Relatividad General

Capítulo 9

El Principio de Equivalencia

Pero yo, que he hecho la prueba, te puedo asegurar, Simplicio, que una bala de cañón, pesando cien o doscientas libras, o incluso más, llegará al suelo no adelantando más que un palmo a una bala de mosquete pesando sólo media libra, dado que ambos caen desde una altura de 200 codos.
(G. Galilei, en forma del personaje de Salviati en *Discorsi e Dimostrazioni Matematiche, intorno a due nuove scienze*, 1638)

9.1. Masa gravitatoria y masa inercial

En el Capítulo 2 hemos dicho que el origen de la relatividad especial data del siglo XVII, cuando Galileo Galilei formuló su Principio de la Relatividad. De cierta manera también las raíces de la relatividad general remontan a esa época. Es a partir de allí cuando personas como Galilei y Simon Stevin (1548 - 1620) empezaban a darse cuenta de que la velocidad que alcanzan los objetos en caída libre es independiente de la masa de los objetos, en contraste con la intuición cotidiana y la física aristotélica. Es legendario (pero históricamente falso) el experimento de Galilei, donde dejó caer bolas de distinta masa de la Torre de Pisa para observar que llegan al suelo simultaneamente. En realidad Galilei sacó sus conclusiones sobre la caída libre dejando rodar bolas de planos inclinados y las publicó en su libro *De motu* en 1590. El que sí subió a la torre de la catedral de Delft para dejar caer las masas fue el ingeniero y científico flamenco Simon Stevin, que escribió ya en 1586 en su libro *Weeghconst*:

> "Deja caer dos bolas de plomo, una diez veces más grande y más pesada que la otra, [...] juntas, desde una altura de 30 pies, encima de una tabla de madera [...] y resultará que la más ligera no tardará diez veces más que la pesada en recorrer el camino, sino que caerán tan juntos encima de la tabla que los sonidos de ambas parezcan un solo golpe."

La *universalidad de la caída libre* , el hecho de que la velocidad de objetos en caída libre sea independiente de su masa, es la manifestación de una coincidencia remarcable en la mecánica newtoniana: la igualdad entre la masa inercial $m_{\rm i}$ y la masa gravitatoria $m_{\rm g}$.

En la mecánica newtoniana aparecen dos tipos de masas, en distintos contextos y de distintos tipos. La primera es la *masa inercial* $m_{\rm i}$, que aparece en la segunda ley de Newton como constante de proporcionalidad entre la fuerza $\vec{F}$ ejercida en un cuerpo y su aceleración $\vec{a}$:

$$\vec{F} = m_{\rm i}\, \vec{a}. \tag{9.1}$$

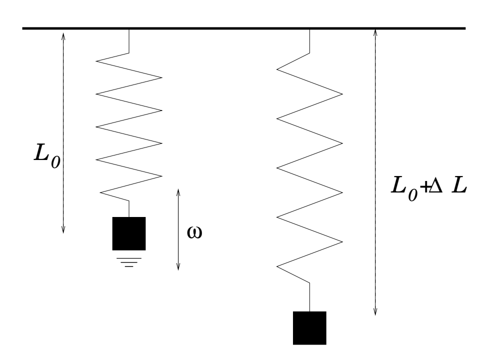


Figura 9.1: *Masa inercial y masa gravitacional: La frecuencia de oscilación ω de un cuerpo colgado de un muelle depende de la masa inercial $m_{\rm i}$ del cuerpo, mientras la extensión ΔL del muelle en la posición de equilibrio depende de la masa gravitacional $m_{\rm g}$.*

La masa inercial es por lo tanto una medida para la resistencia de un cuerpo a cambios de velocidad. Por otra parte, la *masa gravitacional* $m_{\rm g}$ es una medida de cómo interacciona un cuerpo gravitacionalmente con los demás cuerpos en el universo. El potencial gravitatorio causado por un cuerpo sobre otro viene dado por

$$V = \frac{G_N\, m_{\rm g} M_{\rm g}}{r}, \tag{9.2}$$

donde G_N es la constante de Newton y r la distancia entre las dos masas $m_{\rm g}$ y $M_{\rm g}$.

A priori estas dos cantidades físicas no tienen nada que ver la una con la otra y las distintas maneras de medir la masa de un cuerpo miden en realidad una de estas dos posibilidades, dependiendo del experimento. Por ejemplo, colgando una masa de un muelle con constante de elasticidad k, la ecuación de movimiento de la masa viene dada por la segunda ley de Newton

$$kx = m_{\rm i}\ddot{x}, \tag{9.3}$$

de modo que midiendo la frecuencia $\omega = \sqrt{k/m_{\rm i}}$, obtenemos la masa inercial. Sin embargo, en la posición de equilibrio, la fuerza del muelle está compensada por la fuerza gravitacional[1]

$$k\Delta L = m_{\rm g} g, \tag{9.4}$$

donde g es la aceleración gravitatoria y ΔL el desplazamiento de la longitud de reposo del muelle. Por lo tanto, el desplazamiento $\Delta L = m_{\rm g} g/k$ nos da una medida para la masa gravitacional (Véase Figura 9.1).

El hecho de que todos los cuerpos caigan con la misma velocidad implica que hay una relación entre estos dos tipos de masas. Efectivamente, la ecuación de movimiento de una masa en caída libre viene dada por la segunda ley de Newton, donde $\vec{F} = -m_{\rm g} g \vec{e}_z$ es la fuerza de la gravedad galileiana

$$-\, m_{\rm g}\, g \;=\; m_{\rm i}\, \ddot{z}. \tag{9.5}$$

En general la solución de esta ecuación es de la forma

$$z(t) \;=\; -\frac{1}{2}\frac{m_{\rm g}}{m_{\rm i}} g\, t^2 \;+\; v_0 t \;+\; z_0, \tag{9.6}$$

donde v_0 y z_0 son constantes de integración determinadas por las condiciones iniciales. Pero si queremos explicar por qué dos cuerpos con distintas masas (inerciales y gravitacionales) y con

[1]Aquí consideraremos la gravedad galileiana, es decir un campo gravitatorio constante, cerca de la superficie de la Tierra.

las mismas condiciones iniciales llegan a la vez al suelo, tenemos que asumir que

$$\frac{m_{\rm g}}{m_{\rm i}} = \text{constante}. \tag{9.7}$$

En principio podría ser que esta constante dependiera del material, de modo que una bola de acero cayera más rápido o más lento que una bola de madera. Sin embargo, los experimentos enseñan que $m_{\rm g}/m_{\rm i}$ es una constante universal (independiende del material), que se puede tomar 1 con la apropiada elección de unidades.[2]

Efectivamente, en una serie de experimentos históricos se ha comprobado la igualdad de la masa inercial y la masa gravitacional hasta gran precisión. En 1888 el barón húngaro Lóránd von Eötvös (1848 - 1919) inventó una balanza de torsión que midió la diferencia entre el valor de $m_{\rm g}/m_{\rm i}$ de dos cuerpos, utilizando la gravedad y la rotación diaria de la Tierra como medida para $m_{\rm g}$ y $m_{\rm i}$ respectivamente, alcanzando una precisión de $|m_{\rm g} - m_{\rm i}|/m_{\rm g} \lesssim 10^{-9}$. En 1964 el americano Robert Dicke (1916 - 1997) aumentó la precisión del experimento a $|m_{\rm g} - m_{\rm i}|/m_{\rm g} < 10^{-11}$, utilizando la gravedad y la rotación alrededor del Sol. Los límites experimentales actuales son del orden de $3 \cdot 10^{-13}$, mientras hay planes para experimentos en satélites que llegarían a una precisión de 10^{-18}. Veremos ahora como el Principio de Equivalencia explica esta igualdad de masa inercial y gravitatoria de una manera muy natural.

9.2. El Principio de Equivalencia para campos homogéneos

Ya hemos dicho en la sección 5.4 que Einstein ya en 1907 se dio cuenta de que la teoría de la relatividad especial era necesariamente incompatible con la gravedad newtoniana, por ser ésta última una interacción inmediata y a distancia. Sin embargo no estaba claro cómo se podía incorporar la gravedad en la relatividad especial, puesto que intrínsicamente trataba con observadores acelerados, por los cuales el Principio de la Relatividad no es válido.

Por otro lado, a Einstein le impresionó mucho la igualdad de la masa inercial y la masa gravitatoria. Aunque la mecánica newtoniana no daba ninguna razón a priori por la que estas dos cantidades físicas tuvieran el mismo valor, Einstein creía que esto no podía ser casualidad, sino que insinuaba una profunda propiedad de la Naturaleza.

La luz vino en 1907, en lo que Einstein mismo llamó *"la idea más feliz de mi vida"*, cuando se dio cuenta de que un observador en caída libre no siente su propio peso y por lo tanto podría pensar que estuviera en una región del espacio donde no hubiera campo gravitatorio. Efectivamente, una bola, soltada por el observador durante la caída libre, caerá debido a la universalidad de la caída libre a la misma velocidad que el propio observador, de modo que este último verá la bola flotando como si estuviera en el espacio interestelar (véase Figura 9.2).[3] Por lo tanto, un observador en caída libre en un campo gravitacional constante, aunque no es un observador inercial en el sentido newtoniano, podría considerarse a sí mismo en reposo en ausencia de gravedad.

Principio de Equivalencia (para campos gravitatorios constantes): *Un observador en caída libre en un campo gravitatorio constante es equivalente a un observador inercial en ausencia de gravedad. Es imposible determinar la diferencia entre estas dos situaciones a base de experimentos físicos.*

[2]En la segunda mitad de los 1980 parecía haber indicios (aunque contradictorios) de que podría existir una *quinta fuerza* fundamental, aparte de la gravedad, el electromagnetismo y las interacciones fuerte y débil. La quinta fuerza supuestamente actuaría de modo repulsivo a una escala entre 100 y 1000 metros y de manera diferente en objetos de materiales diferentes. Un análisis profundo de los experimentos, sin embargo, ha mostrado que en realidad no existe tal fuerza.

[3]Nótese que esto es en realidad la razón por la que un astronauta no tiene peso en una órbita alrededor de la Tierra: porque está continuamente en caída libre alrededor de la Tierra.

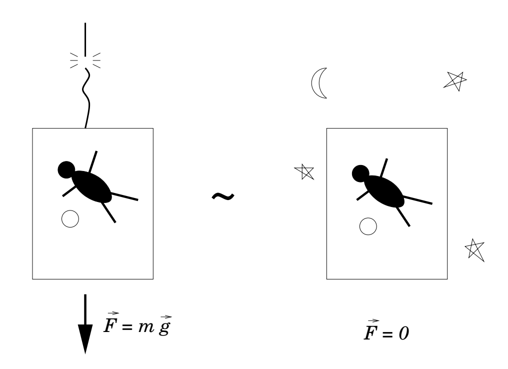


Figura 9.2: *Un observador en caída libre es (localmente) equivalente a un observador en ausencia de campo gravitatorio. No hay un experimento físico que pueda distinguir (¡localmente!) entre estos dos casos.*

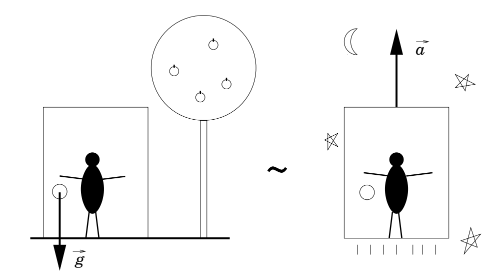


Figura 9.3: *Un observador en un ascensor acelerado con aceleración constante es equivalente a un observador en reposo en un campo gravitacional constante. No hay un experimento físico que pueda distinguir (¡localmente!) entre estos dos casos.*

Por otro lado, un observador $\mathcal{O}'$ en un ascensor que está acelerado en el espacio interestelar con una aceleración constante a verá los mismos efectos que vería un observador en un campo gravitacional constante. En particular, si $\mathcal{O}'$ suelta una bola desde un punto x_0 en el momento t, cuando el ascensor tiene una velocidad v (con respecto a un observador externo $\mathcal{O}$), la bola seguirá por inercia una trayectoria $x(t) = x_0 + vt$ (con respecto a $\mathcal{O}$), dado que el ascensor ya no ejerce ninguna fuerza sobre ella. Sin embargo, el ascensor seguirá acelerando, de modo que la trayectoria de éste viene dada por $X(t) = vt + \frac{1}{2}at^2$ en las coordenadas de $\mathcal{O}$. Mientras las velocidades relativas entre $\mathcal{O}$ y $\mathcal{O}'$ son pequeñas (y por lo tanto $t \approx t'$), vemos que la trayectoria de la bola en las coordenadas de $\mathcal{O}'$ viene dada por

$$x'(t) \;=\; x(t) - X(t) \;=\; x_0 - \frac{1}{2}at^2. \tag{9.8}$$

En otras palabras, para $\mathcal{O}'$ es como si la bola cayera en un campo gravitatorio con aceleración gravitacional $\vec{g} = -\vec{a}$ (véase Figura 9.3). Por lo tanto, un observador moviéndose con aceleración constante, aunque no sea un observador inercial en el sentido newtoniano, se podría considerar en reposo en un campo gravitatorio.

Principio de Equivalencia (para aceleraciones constantes): *Un observador en movimiento uniformamente acelerado es equivalente a un observador inercial en un campo gravitatorio constante. Es imposible determinar la diferencia entre estas dos situaciones a base de experimentos físicos.*

Nótese que para $\mathcal{O}$ la bola choca con el suelo del ascensor porque el suelo alcanza la bola, al no estar acelerada ésta última, mientras $\mathcal{O}'$ explica el mismo fenómeno diciendo que la bola siente

un campo gravitatorio constante. Lo que para $\mathcal{O}$ es inercia, para $\mathcal{O}'$ es gravedad. Esto explica la universalidad de la caída libre y el hecho de que no haya distinción entre masa inercial y masa gravitatoria: dos bolas en caída, aunque tengan masas distintas o estén hechas de materiales distintos, alcanzarán la misma velocidad, puesto que siempre se podría considerar que en lugar de estar en el campo gravitatorio de la Tierra, que el suelo se esté acelerando con una aceleración $a = 9,81 m/s^2$ hacia arriba.

En realidad también se puede dar la vuelta al Principio de Equivalencia y decir que un campo gravitatorio constante en realidad no es más que un artefacto de una elección de unas coordenadas específicas. Efectivamente, la ecuación de movimiento de una partícula en caída libre en un campo gravitatorio constante viene dada por la segunda ley de Newton

$$\ddot{x} = -g. \tag{9.9}$$

Sin embargo, por el Principio de Equivalencia podríamos tomar el punto de vista de un observador en caída libre a través del cambio de coordenadas

$$y \;=\; x \;+\; \frac{1}{2} g t^2, \tag{9.10}$$

de modo que en las coordenadas nuevas la ecuación de movimiento de la partícula viene dada por

$$\ddot{y} \;=\; 0. \tag{9.11}$$

Vemos por lo tanto que en las coordenadas y la ecuación de movimiento corresponde a la de una partícula libre, aunque no en coordenadas cartesianas, sino en coordenadas curvilíneas. Aún así es remarcable que hemos podido eliminar el campo gravitatorio (constante!) a través de un cambio de coordenadas. En la mecánica clásica, las fuerzas que dependen del sistema de referencia y que desaparecen con los cambios de coordenadas, se suelen llamar *fuerzas ficticias*, como la fuerza centrífuga o la fuerza de Coriolis. Lo que nos está diciendo el Principio de Equivalencia es que la fuerza gravitatoria generada por un campo gravitatorio constante no es una fuerza física, sino un artefacto del sistema de referencia.

El hecho de que observadores en caída libre en un campo gravitatorio constante son indistinguibles de observadores inerciales es de suma importancia para compaginar la interacción gravitatoria con la relatividad especial. Recuérdese que la dificultad de incorporar la gravedad en la relatividad especial era que ésta última solamente era válidad para observadores inerciales, mientras que los campos gravitatorios necesariamente implicaba la presencia de observadores acelerados. El Principio de Equivalencia por lo tanto resuelve este problema, por lo menos en el caso de campos gravitatorios constantes, al permitir que se considera éstos como inerciales.

Veremos en la sección 9.4 que la situación es bastante más complicada para los campos gravitatorios inhomogéneos, pero estudiemos primero unas consecuencias directas del Principio de Equivalencia en campos gravitatorios constantes.

9.3. Experimentos en un ascensor

El Principio de Equivalencia, en su forma sencilla, para campos y aceleraciones constantes, llevado a sus últimas consecuencias, predice unos efectos relativistas remarcables. Estudiaremos estos efectos en detalle y con la necesaria precisión matemática en el Capítulo 11, pero aquí tomaremos un primer contacto, partiendo de primeros principios.

Volvamos a la comparación de un observador en un campo gravitatorio constante y un observador en movimiento uniformemente acelerado. El Principio de Equivalencia dice entonces que no existe ningún experimento físico que sea capaz de distinguir entre los estados de movimiento de los dos observadores. En todo esta sección también supondremos que todas las velocidades

v y diferencias de velocidades Δv sean pequeñas en comparación con la velocidad de la luz, de modo que en primera aproximación no nos tenemos que preocupar por las correcciones debido a efectos de la relatividad especial. Para un análisis más profundo referimos a la literatura.

Concretamente, consideramos un observador $\mathcal{O}'$ en un ascensor que se mueve con una aceleración constante a con respecto a un observador externo $\mathcal{O}$. Los dos observadores $\mathcal{O}$ y $\mathcal{O}'$ estarán de acuerdo en que hay una aceleración relativa entre ellos, pero no se pondrán de acuerdo sobre quién se mueve y quién está en reposo, ya que $\mathcal{O}'$ puede insistir en que él no está en movimiento, sino en reposo en el campo gravitatorio con aceleración gravitacional $g = -a$ y que la aceleración relativa entre los dos es debida a que $\mathcal{O}$ está en caída libre en ese campo gravitatorio. Esta es la razón, según $\mathcal{O}'$, por la que $\mathcal{O}$ no nota la fuerza del supuesto campo gravitatorio: precisamente porque está en caída libre.

Para poder mantener la afirmación de $\mathcal{O}'$ de que está en reposo en un campo gravitatorio, la física que vea $\mathcal{O}'$ y que vea un observador $\mathcal{O}''$ en un campo gravitatorio constante, tiene que ser la misma, en el sentido de que cualquier fenómeno que ve uno de ellos, también lo tiene que notar el otro. Examinaremos algunos de estos fenómenos.

Ya lo hemos dicho en la sección anterior, pero merece la pena volver a mencionarlo aquí. Si el observador $\mathcal{O}'$ en el ascensor suelta un masa en el momento $t = t_0$ cuando el ascendor tiene una velocidad $v(t_0)$ relativo a $\mathcal{O}$, esta masa seguirá moviéndose con la velocidad $v(t_0)$, independientemente del ascensor. Por lo tanto el observador externo $\mathcal{O}$ verá que el fondo del ascensor se acerca al objeto soltado con una velocidad relativa $v(t) = v(t_0) + at$ y se choca con la masa después de un tiempo $t = \sqrt{2a/h}$, donde h es la altura dentro del ascensor desde la cual $\mathcal{O}'$ ha soltado el objeto. El observador $\mathcal{O}'$ verá el mismo fenómeno, pero lo interpretará de manera distinta. Para él, la masa se mueve de manera acelerada hacia abajo, debido a la fuerza gravitatoria. El tiempo que tardan distintas masas en llegar al suelo y las velocidades que alcanzan son siempre los mismos. Para $\mathcal{O}$ esto está claro a base de la inercia, ya que una vez soltadas, ya no actúa ninguna fuerza sobre ellas. Por otro lado $\mathcal{O}'$ concluye que el tiempo de caída y las velocidades alcanzadas son independientes de la masa y la composición de los objetos soltados, precisamente lo que observa también el observador $\mathcal{O}''$ en un campo gravitacional. Esto no sólo explica la universalidad de la caída libre, sino también indica que *la gravedad se acopla de manera universal a toda la materia*, independientemente de la masa o la composición. No hay manera de que una partícula no note la gravedad, en contraste con el electromagnetismo, donde una partícula neutra no interacciona con los campos electromagnéticos.

No sólo partículas masivas están afectadas por la gravedad, sino también la luz. Si el observador $\mathcal{O}$ emite una señal de luz que entra horizontalmente por el lateral del ascendor a una altura h, saldrá la luz por el otro lateral a una altura $h - \frac{1}{2}at^2$, donde t es el tiempo que necesita la señal en atravesar el ascensor. Para $\mathcal{O}$ la trayectoria de la luz es una recta, ya que es el ascensor el que se mueve hacia arriba. Sin embargo el observador $\mathcal{O}'$ verá en sus propias coordenadas que la luz sigue una trayectoria curva, que él atribuirá a la presencia de campo gravitatorio. Una primera predicción por lo tanto de la relatividad general, o mejor dicho del Principio de Equivalencia, es que hay una desviación de la luz de su trayectoria recta en presencia de gravedad. Efectivamente un observador $\mathcal{O}''$ en un campo gravitatorio ve la luz desviada si pasa cerca de cuerpos masivos, como fue comprobado exprimementalmente en el famoso eclipse solar de 1919.[4] El mismo fenómeno es responsable de los efectos de las lentes gravitatorias y las cruces de Einstein observadas en imagenes de galaxias lejanas. Discutiremos esto en más detalle en la sección 11.3.

Otro experimento que puede hacer el observador $\mathcal{O}'$ en el ascensor es emitir una serie de pul-

[4]Einstein hizó una primera estimación del efecto de la desviación de la luz cerca del sol en 1911, basado puramente en el Principio de Equivalencia, obteniendo una deflexión de 0,83 arcosegundos. Se hicieron intentos de medir el efecto en eclipses solares en Brasil en 1912 y en Rusia en 1914, pero fracasaron debido al mal tiempo y el estallido de la Primera Guerra Mundial. En 1915 Einstein volvió a calcular el efecto, ahora en base a las ecuaciones de Einstein completas y se dio cuenta de que el resultado verdadero es una desviación de 1,7 arcosegundos, el doble de su primera estimación. Fue este resultado el que fue confirmado en el eclipse de 1919 y que dio a Einstein su fama mundial. Uno sólo puede adivinar lo que habría pasado si las expediciones de 1912 y 1914 no hubieran tenido la mala suerte que tuvieron.

sos periódicos con perido T_e desde el fondo del ascensor hacia el techo. El observador exterior $\mathcal{O}$ verá que en el tiempo Δt entre la emisión en el fondo y la detección en el techo, el ascensor ha adquerido una velocidad adicional $\Delta v = a\Delta t$, de modo que hay una velocidad relativa entre la fuente en el momento de emisión y el detector en el momento de detección, lo que da lugar a un efecto Doppler: el detector verá llegar los pulsos con una frecuencia más baja que la frecuencia de emisión. Este mismo efecto lo verá también $\mathcal{O}'$, pero él, creyendo que está en un campo gravitacional, dirá que el efecto Doppler es debido al hecho de que la luz pierde energía al salir el pozo potencial. En 1960 y 1964 Robert Pound y Glen Rebka hicieron una verificación experimental en un campo gravitatorio real, confirmando que efectivamente un *efecto Doppler gravitacional* existe, tal como lo predice el Principio de Equivalencia. La luz que escapa de un campo gravitatorio sufre un corremiento hacia el rojo, mientras la luz que entra en un pozo potencial gravitacional un corremiento hacia el azul. Discutiremos esto en más detalle en la sección 11.4.

Se puede hacer fácilmente una estimación del efecto en primera aproximación. El emisor emite la primera señal en el momento $t = 0$ y llega en el momento $t = \Delta t$,[5] de modo que la señal en este tiempo ha viajado una distancia

$$\Delta t = h + \frac{1}{2}a(\Delta t)^2, \tag{9.12}$$

donde h es la altura del ascensor, es decir, la distancia entre el emisor y el detector en reposo. De (9.12) podemos calcular una expresión para Δt en función de h y a:

$$\Delta t \;=\; \frac{1}{a}\Big(1 - \sqrt{1-2ah}\Big) \;\approx\; h, \tag{9.13}$$

donde en la última igualdad hemos supuesto que $ah \ll 1$ y desarrollado la raíz cuadrada en un desarrollo de Taylor, quedándonos sólo con el primer orden.[6] Físicamente esto corresponde al hecho de que la aceleración es suficientemente pequeña para que el cambio de velocidad Δv en el intervalo de tiempo Δt sea despreciable. En el momento $t = T_e$ el emisor emite la segunda señal, que llega al detector en el momento $\Delta t + T_d$. (En otras palabras, T_e es la periodo de la señal en el momento de emisión y T_d el periodo en el momento de detección.) La distancia recorrida por la segunda señal es por lo tanto

$$\begin{aligned} \Delta t + T_d - T_e &= h + \frac{1}{2}a(\Delta t + T_d)^2 - \frac{1}{2}aT_e^2 \\ &\approx h + \frac{1}{2}a(\Delta t)^2 + aT_d\Delta t, \end{aligned} \tag{9.14}$$

donde en la última igualdad hemos supuesto que T_d y T_e son pequeñas en comparación con Δt. Podemos obtener fácilmente la relación entre T_e y T_d restando (9.12) de (9.14),

$$T_d - T_e \approx aT_d\Delta t, \tag{9.15}$$

o sustituyendo la expresión (9.13) para Δt

$$T_e \approx \Big(1 - ah\Big)T_d, \tag{9.16}$$

de modo que el periodo de la señal es mayor en el momento de detección que en el momento de emisión. En otras palabras la señal ha sufrido un corrimiento hacia el rojo con un factor $(1 - ah)$. Dado que para el observador $\mathcal{O}'$, que se cree en un campo gravitatorio, el factor ah es la diferencia

[5] Estrictamente hablando, el tiempo t corresponde al tiempo del observador externo $\mathcal{O}$, pero para velocidades pequeñas en comparación con la velocidad de la luz, la diferencia entre el tiempo de $\mathcal{O}'$ y $\mathcal{O}$ es depreciable.

[6] Obviamente, la ecuación (9.12) tiene otra solución, pero ésta es negativa, de modo que se puede rechazar por razones físicas.

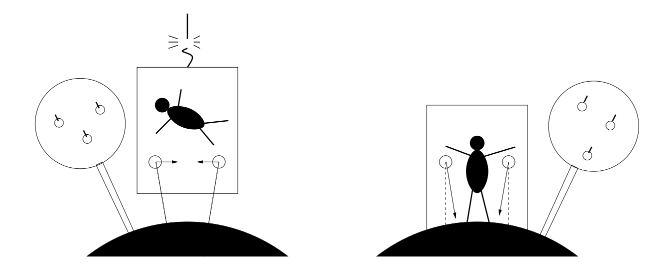


Figura 9.4: *Los efectos de marea: el Principio de Equivalencia sólo es valido localmente, a escalas pequeñas. A escalas más grandes se puede ver la diferencia entre gravedad y aceleración a través de las fuerzas de marea, debido a la inhomogeneidad del campo gravitacional.*

del potencial gravitatorio $\Phi(x)$ entre el emisor y el detector, el Principio de Equivalencia predice que en un campo gravitatorio constante la luz sufre un efecto Doppler gravitacional

$$T_e \approx \Big[1 - \Big(\Phi(h) - \Phi(0)\Big)\Big] T_d. \tag{9.17}$$

En la sección 11.4 deduciremos una expresión más exacta para el efecto Doppler gravitacional en un campo inhomogéneo.

Por último, el efecto Doppler gravitacional tiene una consecuencia directa: la *dilatación gravitacional del tiempo*. El observador $\mathcal{O}'$ podría considerar al emisor como un reloj que emite pulsos con una cierta frecuencia y comparar esta frecuencia con la de un reloj similar situado donde el detector. Debido al efecto Doppler gravitacional, notará que el reloj en el suelo del ascensor corre menos rápido que el reloj en el techo. Esto se corresponde con el hecho de que un reloj abajo en un campo gravitatorio va más lento que uno más arriba, donde la diferencia viene dada por (9.17).

9.4. El Principio de Equivalencia para campos inhomogéneos

Hasta ahora hemos tenido mucho cuidado al considerar el Principio de Equivalencia solamente en el caso de campos gravitatorios constantes o aceleraciones constantes. Sin embargo, en la Naturaleza los campos gravitatorios son causados por la presencia de materia y cada distribución de materia da lugar a una configuración específica de campo gravitatorio, en general inhomogénea. Por lo tanto un campo gravitatorio constante es en realidad una aproximación a pequeña escala, válida en una región pequeña cerca de la superficie de objetos grandes, como la Tierra y habrá que estudiar qué es lo que queda del Principio de Equivalencia para campos gravitatorios generales.

El problema que encontramos es que el Principio de Equivalencia deja de ser válido para campos gravitatorios generales, ya que precisamente las inhomogeneidades del campo hacen que un observador puede distinguir entre estar en caída libre o ser un observador inercial en ausencia de gravedad. Por ejemplo, en un campo gravitatoria radial (producido por una masa esféricamente simétrica, como la Tierra) dos bolas en caída libre no seguirán trayectorias paralelas, sino que se acercarán lentamente según vayan cayendo cada vez más hacía el centro del campo gravitatorio. Observando con suficiente precisión, un observador en caída libre puede detectar este efecto y darse cuenta de que se encuentra en un campo gravitatorio real (véase Figura 9.4).

El carácter inhomogéneo del campo gravitatorio también nos impide deshacer los efectos de la gravedad a través de un cambio de coordenadas. Al aplicar en cambio de coordenadas (9.10),

$$y = x + \frac{1}{2}\, g(x_0) t^2, \tag{9.18}$$

en un punto p (con coordenadas x_0) de un campo gravitatorio inhomogéneo, uno puede conseguir la ecuación de movimiento tome la forma (9.11) en ese punto p, pero no en un punto q cercano, ya que el valor de la aceleración gravitatoria $g(x)$ varía de punto en punto. En contraste con los campos gravitatorios homogéneos, uno no puede considerar la fuerza gravitatoria inhomogénea como una fuerza ficticia.

El hecho de que el Principio de Equivalencia no sea válido en general, pone en serio peligro nuestra estrategia para compaginar los observadores acelerados en campos gravitatorios con el Principio de la Relatividad. Concretamente, los efectos causados por las inhomogeneidades de un campo gravitatorio general se llaman *fuerzas de marea* y son justo estos efectos los que hacen que un observador sea capaz de distinguir entre un sistema inercial y un sistema en caída libre. Sin embargo, estas fuerzas de marea son pequeñas a escalas pequeñas, ya que son proporcionales al gradiente del campo gravitatorio. Por lo tanto, a escalas tan pequeñas que el campo gravitatorio parece localmente constante, las fuerzas de marea desaparecen y el Principio de Equivalencia vuelve a recuperar su validez, aunque sea localmente.

Principio de Equivalencia (formulación general): *Observadores en caída libre en un campo gravitatorio general son localmente equivalentes a observadores inerciales. No hay experimentos locales que puedan distinguir entre estas dos situaciones.*

Es importante darse cuenta del validez local, no global, del Principio de Equivalencia: considera dos observadores en caída libre cerca de la Tierra, uno encima del ecuador y otro encima del Polo Norte. Cada uno de estos observadores puede considerarse (brevemente) en resposo en ausencia de gravedad, pero aún así verá que el otro observador está acelerado con respecto a él. No existe un sistema de referencia donde ambos observadores están simultáneamente en reposo, por la sencilla razón de que las inhomogeneidades del campo gravitatorio causa aceleraciones relativas entre los dos.

La pregunta que entonces surge es ¿cuál es esa escala a partir de la que el Principio de Equivalencia pierde su validez? ¿A qué se refiere exactamente con experimentos locales? Básicamente el Principio de Equivalencia es válido mientras que los efectos de marea sean indetectables. Ya hemos dicho que los efectos de marea son proporcionales al gradiente del campo gravitacional: los efectos serán más grandes cerca de una masa puntual que lejos de ella. Pero el gradiente del campo gravitacional no es el único parámetro: el rango de validez del Principio de Equivalencia también depende de la precisión de los experimentos a nuestra disposición. Si se tiene unos aparatos con un error experimental ΔL, se verá que el Principio de Equivalencia se cumple en regiones del espaciotiempo donde el gradiente del campo gravitatorio no es mayor que $1/\Delta L$.[7] Al obtener aparatos nuevos con una precisión más alta, la región de validez se reducirá correspondientemente.

A primera vista esto parece reducir el Principio de Equivalencia a una tautología: en regiones donde no soy capaz de medir nada, no veré ningún efecto. Sin embargo, esto no es cierto. Uno podría imaginarse perfectamente una teoría donde en regiones donde los experimentos no detectan ningún gradiente apreciable en el campo graviatatorio, objetos de distinta masa y/o distinta composición cayesen a velocidades distintas. De hecho, esto es lo que pasa con la fuerza electromagnética: objetos con la misma carga pero con masas distintas, como el protón y el nucleo de deuterio ($^2H^+$), aceleran de manera distinta en un campo eléctrico. Ambas partículas sienten la misma fuerza atractiva por tener la misma carga eléctrica, pero la inercia del deuterio es dos veces mayor por tener el doble de masa. Sin embargo en el caso de la gravedad, la "carga gravitatoria" es la misma masa que causa su inercia, de modo que la aceleración gravitatoria de todas

[7]Con regiones pequeñas del espaciotiempo queremos decir no solamente regiones espaciales pequeñas (i.e. ascensores de tamaño limitado), sino también intervalos de tiempo pequeños en comparación con el tiempo que necesita la luz para recorrer una distancia en que el gradiente llegue a ser apreciable. Un efecto de marea pequeño podría acumularse en un intervalo de tiempo grande y llegar a ser notable de esta manera.

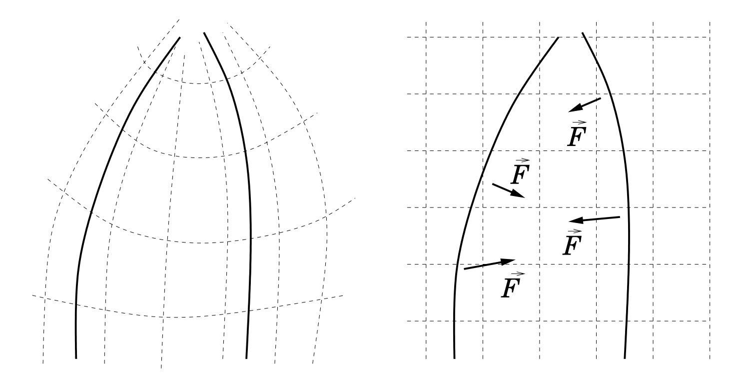


Figura 9.5: *La gravedad como manifestación de la curvatura del espacio. Las trayectorias de partículas libres en un espacio curvo son las geodésicas de la variedad. Sin embargo un observador que insiste en utilizar las coordenadas de un espacio plano se ve obligado a invocar fuerzas a distancia $\vec{F}$ para describir las aceleraciones observadas.*

las partículas es la misma, independientemente de su masa. Esto es lo que hace que el Principio de Equivalencia no es trivial, sino muy profundo.

9.5. Implicaciones matemáticas del Principio de Equivalencia

La validez del Principio de Equivalencia, y precisamente su carácter local, tiene importantes implicaciones sobre la estructura matemática del espaciotiempo. El hecho de que localmente un observador no pueda distinguir entre estar en caída libre en un campo graviatatorio o ser un observador inercial interestelar implica que existe un cambio de coordenadas que elimina el campo gravitatorio en una pequeña región y hace que el espaciotiempo parezca localemente Minkowski. Debido a las inhomogeneidades del campo gravitatorio, ese cambio de coordenadas será distinto en cado punto del espaciotiempo y en general no será posible encontrar una transformación tal que el espacio entero se reduzca al espacio de Minkowski. La imagen que surge por lo tanto es la de un espacio que en cada punto es localmente plano (Minkowski), pero globalemente no. Esto es el concepto de una variedad con curvatura, que encontramos en la Parte II.

Efectivamente, el espaciotiempo de la relatividad general es una variedad lorentziana cuadrimensional con curvatura determinada por la conexión de Levi-Civita, donde las partículas libres (es decir, las partículas sobre las que únicamente actúa la gravedad) viajan a lo largo de las geodésicas de la variedad curva. En particular, el hecho de que el espaciotiempo tenga una curvatura no-trivial hace que las partículas libres sufren un efecto de desviación geodésica: partículas que inicialmente siguen trayectorias paralelas notarán aceleraciones relativas entre ellas.

Sin embargo, si uno insiste en que el espacio es plano y descriptible en términos de coordenadas cartesianas, uno tiende a invocar la Segunda Ley de Newton y adscribir estas aceleraciones a alguna fuerza $\vec{F}$, que actúa a distancia y se transmite por el vacío (véase Figura 9.5). La interpretación relativista, por lo tanto, de lo que en la mecánica newtoniana es la fuerza gravitatoria $\vec{F}_{\rm grav}$ es justamente esto: una pseudo-fuerza debido al uso de un sistema de coordenadas "inapropiado". En realidad, la interacción gravitatoria no es más que una manifestación de la curvatura del espaciotiempo. La gravedad es geometría.

La mejor ilustración de la diferencia entre estas dos maneras de pensar es la interpretación de la ecuación de movimiento de una partícula en un campo gravitatorio. Donde para Newton la

partícula obedece la segunda ley de Newton

$$\ddot{x}^i - \frac{F^i_{\text{grav}}}{m} = 0, \tag{9.19}$$

para Einstein, la partícula sigue una geodésica en el espaciotiempo curvo,

$$\ddot{x}^\mu + \Gamma^\mu_{\nu\rho}\dot{x}^\nu \dot{x}^\rho = 0. \tag{9.20}$$

En otras palabras, lo que para Newton es la fuerza gravitatoria a distancia $\vec{F}_{\text{grav}}/m$, es para Einstein un término puramente geométrico, $\Gamma^\mu_{\nu\rho}\dot{x}^\nu \dot{x}^\rho$, debido a la curvatura no-trivial del espaciotiempo.

En la sección 8.5 hemos visto que alrededor de cualquier punto de una variedad se puede considerar una pequeña región en la cual el espacio tangente era una buena aproximación para la variedad. Además en esta pequeña región existen unas coordenadas, las llamadas coordenadas localmente inerciales, tales que en una región pequeña alrededor de este punto los símbolos de Christoffel son cero y la métrica se reduce a la plana en primera aproximación. Esta propiedad de las variedades de es el fundamento matemático del Principio de Equivalencia: el observador que se cree localmente inercial es el observador que usa las coordenadas localmente inerciales y ha eliminado el campo gravitatorio, por lo menos en una pequeña región alrededor de sí mismo.

El hecho de que el cambio de coordenadas (8.48) entre unas coordenadas arbitrarias x^μ y las coordenadas localmente inerciales y^α sea un cambio general de coordenadas, implica que las leyes de la física tienen que transformar bien bajo un grupo más grande que el grupo de Lorentz. Efectivamente, en las coordenadas localmente inerciales, un observador tiene todo el derecho a llamarse un observador inercial y por lo tanto para él las leyes de la físcia tendrán la forma de las leyes de la relatividad especial, por lo menos localmente. Sin embargo, el mismo observador podría estudiar los fenómenos físicos en cualquier otro sistema de coordenadas y debería llegar a las mismas leyes de la física, aunque escritas en coordenadas distintas. Esta observación nos lleva a otro principio básico:

Principio de Covariancia: *Las leyes de la física deben tener la misma forma en todos los sistemas de referencia. Las leyes de la física deben por lo tanto transformar de manera covariante bajo cambios generales de coordenadas.*

En la práctica esto implica que una ley es válida en general si es válida en relatividad especial y si la escribimos de manera covariante, es decir, en función de objectos que transformen bien bajo cambios generales de coordenadas. En particular, sustituyendo derivadas parciales por derivadas covariantes y la métrica $\eta_{\mu\nu}$ por una métrica general $g_{\mu\nu}$. Einstein llegó a esta conclusión en 1913, más de 5 años después del Principio de Equivalencia y aún así no lo utilizó en sus primeras formulaciones de la relatividad general, porque temía que no diese el límite newtoniano correcto y violara la causalidad y el determinismo.[8] Sólo en 1915 llegó a una versión de las ecuaciones de Einstein que estaba de acuerdo con ambos principios.

Vemos por la tanto que el Principio de Equivalencia y la geometría diferencial son dos caras de la misma moneda, que expresan, cada uno de su manera, física o matemática, el carácter del espaciotiempo como variedad lorentziana con curvatura. No nos sorprenderá por lo tanto que todas las herramientas de la geometría diferencial que hemos visto en la Parte II nos van a resultar extremadamente útiles para toda la física gravitacional relativista.

[8]En particular le preocupaba que unas ecuaciones covariantes no determinan por completo la forma de la métrica a partir de unas condiciones iniciales y tardó más de dos años en comprender que los grados de libertad no determinados no eran físicos, sino que corresponden a la libertad de elección de coordenadas.

9.6. Curvatura, sensación de peso y sistemas inerciales

En las secciones anteriores hemos encontrado diferentes tipos de observadores: observadores acelerados y no-acelerados, observadores en el espacio plano y el espacio curvo, observadores que notaban o no un campo gravitatorio y hasta observadores acelerados que se creían inerciales y vice versa. Hemos visto el Principio de Equivalencia adscribe (por lo menos localmente) las mismas sensaciones y resultados experimentales a observadores que claramente están en situaciones muy distintas y hemos encontrado que en un caso la fuerza gravitatoria es una manifestación de la curvatura del espaciotiempo y en otro caso un artefacto del uso de coordenadas curvilineas. Aunque es extremadamente importante la lección que nos da el Principio de Equivalencia sobre el carácter del campo gravitatorio, es fácil perderse entre las distintas posibilidades y tener la sensación de que los árboles no le dejan a uno ver el bosque.

Una primera fuente de posible confusión es el uso poco riguroso del término "gravedad". Hasta ahora hemos usado "fuerza gravitatoria" para referirnos a dos cosas que en realidad son muy distintas: la curvatura del espaciotiempo y la sensación de tener peso. Ambos conceptos representan ciertos aspectos de lo que solemos denominar con el término general "gravedad", pero cuya distinción es fundamental para entender bien las implicaciones del Principio de Equivalencia. El hecho de que son dos cosas distintas se ve en que algunos observadores notan uno de estos aspectos, otros el otro, algunos notan los dos y otros ninguno.

La gravedad como sensación de peso es lo que intuitivamente entendemos como la fuerza gravitatoria (a la Newton) y es lo que más se acerca a la idea cotidiana de gravedad. Es la sensación de tener que luchar contra una fuerza permanente y de cansarse si uno está mucho rato de pie. Es lo que permite distinguir entre arriba y abajo y hace que los objetos caen hacia abajo. La ausencia de la sensación de peso es el cosquilleo que uno siente en la tripa en una montaña rusa o al caerse de una cierta altura, debido al hecho de que en ese momento las vísceras ya no pesan en la cadera. La sensación de peso es un concepto muy físico (fisiológico, en contraste con matemático) y depende de la manera en que se mueva un observador. Por esa razón es posible aumentar esta sensación de peso o hacerla desaparecer al cambiar el estado de movimiento, es decir con un cambio de coordenadas: dos observadores moviéndose de manera distinta en el mismo punto del espacio pueden llegar a tener sensaciones de peso completamente distintas.

Por otro lado, gravedad como manifestación de la curvatura del espacio es un concepto mucho más objetivo (hasta se puede decir que es un concepto absoluto), ya que describe las propiedades matemáticas del espaciotiempo considerado. La curvatura del espaciotiempo es la responsable de que el resultado de transporte paralelo de un vector depende de la curva seguida y que geodésicas cercanas sufren una desviación geodésica. La ausencia de curvatura hace que todos los observadores recuperen los resultados de la geometría euclídea, independientemente de cómo se muevan o de las coordenadas que usen. Dos observadores moviéndose de manera distinta en el mismo espacio verán exactamente las mismas propiedades geométricas, ya que éstas con intrínsicas de la variedad considerada, no del sistema de referencia elegido.

Ahora, es importante darse cuenta de que estos dos conceptos son completamente independientes, precisamente porque uno es subjetivo y el otro no. Un observador en el espacio plano puede tener sensación de peso si está acelerado y sentirá ingravidez si no lo está, mientras que un observador en el espacio curvo siente ingravidez si sigue una trayectoria geodésica, pero si no, no. Discutamos esto en un poco más de detalle considerando cuatro observadores distintos y comparando lo que tienen en común y en qué se distinguen entre ellos.

Los cuatro observadores son básicamente los que hemos encontrado en la sección 9.2: dos observadores $\mathcal{O}_1$ y $\mathcal{O}_2$ en el espacio interestelar y dos observadores $\mathcal{O}_3$ y $\mathcal{O}_4$ cerca de la superficie de un planeta masivo (véase Figura 9.6). Supondremos que $\mathcal{O}_1$ y $\mathcal{O}_2$ están tan alejados de cualquier concentración de materia, que se pueden considerar que en una región muy grande alrededor de ellos el espacio es (de modo efectivo) Minkowski. Lo que distingue $\mathcal{O}_1$ y $\mathcal{O}_2$ es que $\mathcal{O}_2$ se mueve con una aceleración constante con respecto al primero. Por otro lado, la masa del planeta de $\mathcal{O}_3$ y

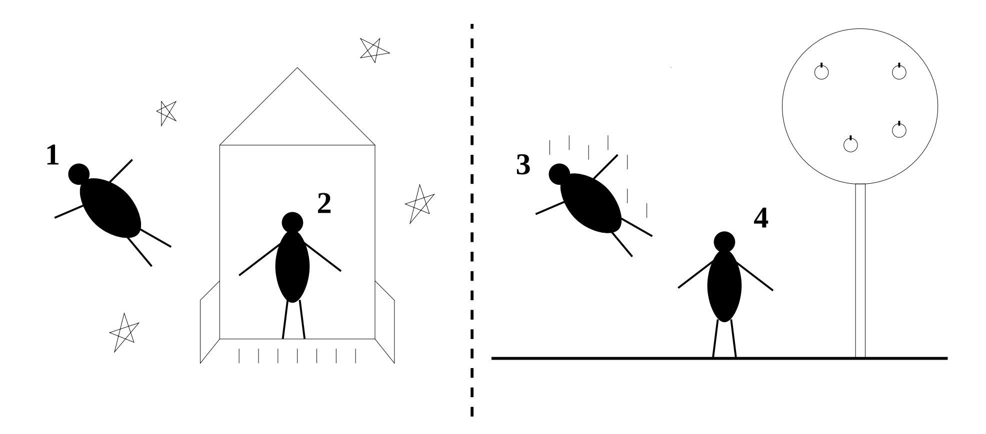


Figura 9.6: *Cuatro observadores inequivalentes: un observador inercial en el espacio plano (1), un observador acelerado en el espacio plano (2), un observador en caída libre en un espacio curvo (3) y un observador en resposo en un espacio curvo (4). Cada observador tiene algo en común con cualquier de los otros tres observadores, pero también algo que les distingue, de modo que los cuatro representan situaciones físicas distintas.*

$\mathcal{O}_4$ curva el espaciotiempo, como veremos en el Capítulo 12, de modo que la geometría que ven éstos no es plana. Dentro de este espaciotiempo curvo, $\mathcal{O}_3$ está en caída libre, mientras que $\mathcal{O}_4$ se queda en reposo en la superficie del planeta (que supondremos estático y sin rotación).

Por lo tanto $\mathcal{O}_1$ y $\mathcal{O}_2$ ven la misma geometría, aunque la midan en coordenadas distintas: $\mathcal{O}_1$ usa de manera natural las coordenadas cartesianas de la relatividad especial, mientras que $\mathcal{O}_2$ utiliza las coordenadas correspondientes a un sistema acelerado, de las que en el Capítulo **??** veremos que se llaman las *coordenadas de Rindler*. Aún así, ambos observadores coincidirán sobre las propiedades geométricas del espaciotiempo y los efectos que éstas tienen sobre la física. Por ejemplo, ambos estarán de acuerdo en que las geodésicas son aquellas curvas que $\mathcal{O}_1$ llama rectas y que dos rayos de luz inicialmente paralelas no se cruzan nunca. De la misma manera $\mathcal{O}_3$ y $\mathcal{O}_4$ coincidirán en que en su espacio hay un efecto de desviación geodésica apreciable, debido a la curvatura causada por la presencia del planeta, a pesar de que $\mathcal{O}_4$ utiliza las coordenadas de Schwarzschild y $\mathcal{O}_3$ unas coordenadas localmente inerciales. De la misma manera estarán de acuerdo en que no es posible sincronizar dos relojes en reposo a distintas alturas sobre el suelo. En otras palabras, $\mathcal{O}_3$ y $\mathcal{O}_4$ dirán ambos que en su espacio hay una campo gravitatorio en el sentido de curvatura, mientras $\mathcal{O}_1$ y $\mathcal{O}_2$ coincidirán que en su caso no lo hay. Por lo tanto, la geometría del espacio y los procesos físicos que ocurren son absolutos e independientes del observador.

Aunque $\mathcal{O}_1$ y $\mathcal{O}_2$ coincidan sobre las propiedades matemáticos de su espacio, la percepción (la sensación fisiológica y la interpretación de los hechos) de cada uno es muy distinta: $\mathcal{O}_1$ está experimentando una ingravidez, mientras que $\mathcal{O}_2$ está notando claramente su peso. Y exactamente lo misma pasa con $\mathcal{O}_3$ y $\mathcal{O}_4$ respectivamente. Aunque estén en espacios distintos, con geometrías distintas, $\mathcal{O}_1$ y $\mathcal{O}_3$ tiene la misma sensación de peso, al igual que lo tienen $\mathcal{O}_2$ y $\mathcal{O}_4$. Es precisamente eso lo que permite el Principio de Equivalencia relacionar estos observadores y declararles equivalentes (aunque solamente localmente, ya que la curvatura en un caso causa efectos de marea que no ve el otro). Por lo tanto, si debido a su ingravidad $\mathcal{O}_1$ y $\mathcal{O}_3$ afirman de que no hay un gravedad, mientras que $\mathcal{O}_2$ y $\mathcal{O}_4$ afirman que sí la hay, no se están refiriendo a gravedad como curvatura, sino a la sensación de peso.

Aunque $\mathcal{O}_3$ ve los mismos fenómenos físicos que $\mathcal{O}_4$, el hecho de que tenga la misma sensación de peso que $\mathcal{O}_1$, hace que su interpretación de estos fenómenos se parece más a la de $\mathcal{O}_1$ que a la de $\mathcal{O}_4$, como hemos argumentado en las secciones anteriores. Y lo mismo ocurre para $\mathcal{O}_2$ y $\mathcal{O}_4$. Por ejemplo, $\mathcal{O}_2$ puede adscribir su peso a la presencia de un campo gravitatorio pa-

	$\mathcal{O}_1$	$\mathcal{O}_2$	$\mathcal{O}_3$	$\mathcal{O}_4$
Peso	No	Sí	No	Sí
Curvatura	No	No	Sí	Sí
Métrica	$\eta_{\mu\nu}$	$g_{\alpha\beta} = \frac{\partial x^\mu}{\partial y^\alpha}\frac{\partial x^\nu}{\partial y^\beta}\eta_{\mu\nu}$	$\eta_{\alpha\beta} + \mathcal{O}(x^2)$	$g_{\mu\nu}$
$\Gamma_{\mu\nu}^\rho$	$= 0$	$\neq 0$	$= 0 + \mathcal{O}(x)$	$\neq 0$
$R_{\mu\nu\rho}{}^\lambda$	$= 0$	$= 0$	$\neq 0$	$\neq 0$

Cuadro 9.1: *Resumen de lo que cada uno de los cuatro observadores tiene en común y lo que les distingue de los demás observadores: $\mathcal{O}_1$ y $\mathcal{O}_3$ tienen la misma sensación de peso (al igual a $\mathcal{O}_2$ y $\mathcal{O}_4$), por lo que el Principio de Equivalencia puede declararles equivalentes. Sin embargo, no ven la misma curvatura, de modo que la equivalencia sólo es local. $\mathcal{O}_1$ y $\mathcal{O}_2$, al estar en el mismo espacio (al igual que $\mathcal{O}_3$ y $\mathcal{O}_4$) ven la misma geometría y los efectos de ésta sobre la física. Nótese que la sensación de peso está codificada en los símbolos de Christoffel, mientras que la curvatura lo está en el tensor de Riemann.*

recido al de $\mathcal{O}_4$, aunque se dará cuenta de su interpretación erronea en cuanto decida cambiar a coordenadas localmente inerciales. En esta caso de encontrará con la sorpresa de que su campo gravitatorio no sólo desaparece en una pequeña región alrededor de él, sino en el espacio entero, ya que sus coordenadas localmente inerciales son en realidad las coordenadas cartesianas de $\mathcal{O}_1$. En este sentido está claro que es $\mathcal{O}_2$ él que está acelerado, y no $\mathcal{O}_1$, ya que sólo $\mathcal{O}_1$ sigue una trayectoria geodésica en el espacio de Minkowski.

Se puede expresar todo este análisis en términos matemáticos: ya hemos visto en la Parte II que la curvatura de un espaciotiempo viene codificada en el tensor de Riemann $R_{\mu\nu\rho}{}^\lambda$, y al ser éste un objeto covariante, todos los observadores del mismo espaciotiempo coincidirán sobre su valor (módulo cambios generales de coordenadas). Por otro lado, la sensación de peso depende de la manera en que se muevan los observadores y se puede eliminar a través de un cambio de coordenadas. Eso es posible porque la sensación de peso está codificada en los símbolos de Christoffel $\Gamma_{\mu\nu}^\rho$, que efectivamente no son tensoriales y pueden ser igualadas a cero localmente a través de un cambio de coordenadas. Además, sabemos que los símbolos de Christoffel pueden ser distintos de cero en coordenadas curvilíneas en el espacio plano, lo que explica por qué el observador $\mathcal{O}_2$ nota su peso aún en ausencia de curvatura, mientras que $\mathcal{O}_3$ no lo nota en un espacio curvo (véase Cuadro 9.1).

Finalmente, nos podemos preguntar a cuál de estos observadores podríamos llamar inerciales. Para la mecánica newtoniana e incluso la relatividad especial, el asunto es muy sencillo: aunque las posiciones y las velocidades son relativas, las aceleraciones son absolutas, en el sentido de que las nota cualquier observador. Por lo tanto un observador puede saber perfectamente si está en un sistema inercial o no: si se nota que las partículas en su sistema de referencia sufren aceleraciones no-inerciales, es decir, aceleraciones que no se atribuyen a fuerzas que actúan sobre las partículas, entonces el sistema no es inercial. Por otra lado, si todas las aceleraciones se pueden adscribir a fuerzas (es decir, si es válida la segunda ley de Newton), el sistema es inercial. En este sentido, Newton no hubiera dudado en llamar inerciales a $\mathcal{O}_1$ y $\mathcal{O}_4$, ya que los dos son observadores estáticos, el primero en ausencia de gravedad y el segundo en un campo gravitatorio, mientras que de ninguna manera hubiera pensado en $\mathcal{O}_2$ y $\mathcal{O}_3$ como inerciales.

Sin embargo, el Principio de Equivalencia complica seriamente esta idea ingenua. Hemos visto que el Principio de Equivalencia relaciona las experiencias de observadores inerciales à la Newton con no-inerciales ($\mathcal{O}_1$ con $\mathcal{O}_3$ y $\mathcal{O}_2$ con $\mathcal{O}_4$), afirmando que por lo menos localmente no

hay distinción entre ellos. En otras palabras, la distinción que era tan obvia para Newton, ya es mucho menos clara, ya que el Principio de Equivalencia implica que no solamente las velocidades, sino también la (sensación de) aceleración es relativa.

Desde el punto de vista de la geometría diferencial se podría definir inerciales como aquellos observadores que siguen geodésicas (los que no tienen sensación de peso, es decir $\mathcal{O}_1$ y $\mathcal{O}_3$), ya que las geodésicas, siendo las trayectorias no-aceleradas un concepto objetivo de la geometría del espacio. Pero aún así esto no encaja con la definición de ingenua de Newton: dado un observador inercial, todos los demás observadores inerciales se tienen que mover de manera uniforme rectilínea con respecto al primero. Sin embargo, dos observadores $\mathcal{O}_3$ y $\mathcal{O}'_3$ cayendo uno sobre el Polo Norte y otro sobre el ecuador tienen una aceleración relativa debido a la desviación geodésica.

Pero si no podemos elegir de manera natural como inercial a un observador específico de una familia de observadores no-acelerados, incluso si los observadores acelerados tienen dificultad para distinguir (localmente) si están acelerado o no, ¿cómo podemos identificar los sistemas inerciales? Es una pregunta importante, ya que uno de los postulados de la relatividad especial es que las leyes de la física son válidas para observadores inerciales. Si no sabemos quiénes son estos observadores inerciales, ¿cómo sabemos para quién la relatividad especial es correcta? En otras palabras, la pregunta se reduce a ¿cómo encaja la relatividad especial dentro del marco de la relatividad general?

En el libro *"The Evolution of Physics"* (1938), Albert Einstein y Leopold Infeld (1898 - 1968) se hacían la pregunta de cómo definir un sistema inercial y, más profundo aún, si es que realmente existen los sistemas inerciales. Su conclusión era que el concepto clásico de sistema inercial no está bien definido, ya que para identificarlo hace falta recurrir a un razonamiento circular: según Newton, un sistema inercial es aquel sistema donde una partícula libre de influencias externas viaja de manera uniforme y rectilínea. Pero la única manera de definir "una partícula libre de influencias externas" es justamente como una partícula que viaja de manera uniforme y rectilínea en un sistema inercial. Einstein e Infeld por lo tanto hacen el comentario irónico

> "Hemos aprendido mucho sobre las leyes de la física: que son invariantes bajo las transformaciones de Lorentz y que son válidas en todos los sistemas inerciales que se mueven de manera uniforme entre ellos. Es cierto que tenemos las leyes, pero no sabemos a qué sistema de referencia se refieren [...] y así parece que toda la construcción mental física está construida sobre arena."

La solución de este problema está en relajar el estatus especial que da la mecánica newtoniana y la relatividad especial a los sistemas inerciales. El Principio de Covariancia dice que el lenguaje tensorial permite escribir las leyes de la física de manera que son válidas para todos los observadores y el Principio de Equivalencia nos enseña que cualquier observador tiene derecho a considerarse inercial, ya que siempre puede interpretar posibles aceleraciones no-inerciales como efectos gravitatorios.

Por lo tanto, también el concepto de sistema inercial se ha vuelto relativo y surge una nueva imagen: en lugar de que exista una única clase de sistemas que es universalmente reconocida como inercial, existen varias clases de observadores que se llamarán inerciales a sí mismos y a los miembros de su clase y no-inerciales a los que no están en su clase. Dos observadores que no pertenecen a la misma clase estarán de acuerdo en que no están en la misma clase, es decir, en que hay aceleraciones relativas entre los dos, pero no estarán de acuerdo sobre quién es el inercial y quién está acelerado. Por ejemplo tanto $\mathcal{O}_3$ como $\mathcal{O}'_3$ y $\mathcal{O}_4$ pueden considerarse inerciales a sí mismos (con peso o sin peso, según el caso) pero verán a cualquier de los otros observadores acelerados.

¿Dónde entra entonces la relatividad especial? Un observador, que se considera a sí mismo inercial y que utiliza coordenadas localmente inerciales, verá el espaciotiempo localmente como Minkowski (en primera aproximación) y dirá que el grupo de simetría es el grupo de Lorentz. Concretamente, está relacionado a través de una transformación de Lorentz con cualquier otro

observador cercano de la misma clase y entre los dos observadores habrá los conocidos efectos relativistas de la sección 3.1, como la dilatación del tiempo y la contracción de Lorentz.

Pero estos efectos sólo serán una primera aproximación a nivel local, ya que el espaciotiempo en coordenadas localmente inerciales sólo se aproxima a Minkowski en una región pequeña. En general el espaciotiempo global no tendrá la simetría del grupo de Lorentz, por la curvatura de la variedad, y dos observadores en distintas regiones del espaciotiempo no están en general relacionados a través de una transformación de Lorentz y se verán mutuamente como acelerados, debido a las inhomogeneidades del campo gravitatorio. En principio incluso no tiene sentido hablar de la velocidad relativa entre estos dos observadores, ya que el sistema de coordenadas localmente inerciales (y por lo tanto el espacio de Minkowski) de cada observador es muy distinto al del otro. Matemáticamente los vectores de velocidad cada observador pertenecen a espacios tangentes distintos y no hay manera natural de compararlos.

En resumen, la relatividad especial está incorporada en la relatividad general como una aproximación a escalas pequeñas, donde es válido el Principio de Equivalencia. Este principio nos dice que cualquier observador se puede considerar localmente como inercial y verá localmente que la relatividad especial es válida.

Capítulo 10

Las ecuaciones de Einstein

El espacio dice cómo se mueve la materia.
La materia dice cómo se curva el espacio.
(J.A. Wheeler)

En la Parte II hemos visto la geometría diferencial como la herramienta matemática necesaria para tratar con espacios curvos y en el Capítulo 10 hemos encontrado argumentos heurísticas para creer que los espacios curvos son una buena descripción de la fuerza gravitatoria. En este capítulo formalizaremos esta idea, dando la formulación exacta de las interacciones gravitatorias en términos de espacios curvos. A partir de ahora asumiremos siempre que la conexión utilizada es la de Levi-Civita.

10.1. El tensor de energía-momento

En el capítulo anterior hemos visto como el Principio de Equivalencia nos da una manera heurística para describir el campo gravitatorio, como la curvatura del espacio. Pero no hemos dicho todavía qué es lo que curva el espacio. De la gravedad newtoniana sabemos que la fuente de la fuerza gravitatoria es la masa de los objetos en el espacio, así que cabe esperar que sea la materia la que curva el espacio. Esto es verdad, aunque no es la única fuente de curvatura.

La relatividad especial nos enseña que la masa de un objeto es sólo una de las posibles manifestaciones de energía y que es posible convertir una cosa en la otra, a través de la relación $E = mc^2$. Por ejemplo, la masa de un nucleo de helio es menor que la suma de las masas de sus componentes (2 protones y 2 neutrones), siendo la diferencia la energía que se libera en un proceso de fusión nuclear (y neutrinos). Para que el Principio de Equivalencia siga siendo válido y la masa gravitacional del nucleo de helio sea igual a la masa inercial, la gravedad se tiene que acoplar no sólo a las masas de los componentes, sino también a la energía (negativa) de ligadura. En general, la gravedad se acopla a cualquier tipo de energía y momento en el espaciotiempo, incluido la energía de cualquier tipo de campos presentes o de la propia curvatura del espacio.

La pregunta ahora surge cómo describir la energía, la masa y el momento de la materia de una manera útil y manejable. En el Capítulo 5 ya hemos dado una descripción para el vector de energía-momento p^μ y la masa m_0 de una partícula, pero esta descripción no es muy práctica si tratamos con grandes cantidades de partículas o con campos. En estos casos es más conveniente describir el sistema como un fluido, un sistema continuo caracterizado por funciones continuas en el espaciotiempo. La gran diferencia es que estas funciones son cantidades macroscópicas, como la velocidad del fluido, la densidad, la presión, etc, más que la energía y el momento de cada

una de las partículas individuales. El objeto que describe la energía y el momento de este fluido es el *tensor de energía-momento* $T^{\mu\nu}$. En esta sección nos limitaremos por razones pedagógicos al caso del espacio plano $g_{\mu\nu} = \eta_{\mu\nu}$. En la sección 10.3 contaremos cómo generalizarlo a espacios arbitarios.

Hay dos maneras de definir el tensor de energía-momento: una manera es a través de un principio variacional de la acción que describe el sistema considerado y esta manera trataremos en la sección 10.3. La otra manera es más directa y más física. La definición entonces del tensor de energía-momento $T^{\mu\nu}$ es el flujo de cuadrimomento p^μ a través de una superficie x^ν constante. Esta definición puede sonar muy poco iluminativa, pero en realidad no es más que la generalización de un vector de corriente, como por ejemplo j^μ.

En electromagnetismo la densidad de corriente $j^\mu = \rho u^\mu$ es el flujo de de carga eléctrica q, una cantidad escalar, a través de una superficie x^μ constante. En particular, la componente j^0, la densidad de carga ρ, se puede ver como el flujo de carga a través de una superficie espacial $t = t_0$. De la misma manera, las componentes espaciales j^i representan el flujo de carga a través de una superficie temporal $x^i = x_0^i$. Además la ley de conservación de carga $\partial_\mu j^\mu = 0$ dice que el cambio de densidad de carga en un volumen es igual al flujo de carga a través la superficie del volumen.

En el caso del tensor de energía-momento, la idea es la misma, sólo que en lugar de ser el flujo de una cantidad escalar, es el flujo de una cantidad vectorial, el cuadrimomento p^μ. El objeto que describe este flujo por lo tanto no es un vector, sino un tensor simétrico $T^{\mu\nu}$ de rango 2.

De este modo, las componentes $T^{\mu 0} = T^{0\mu}$ son el flujo de p^μ a través de una superficie espacial $x^0 = cte$. En particular, T^{00} es la densidad de energía, el flujo de energía en la dirección temporal, mientras T^{0i} es la densidad de momento, el flujo de momento a través de una superficie $x^0 = cte$. Los componentes T^{ij} tienen que ver con las fuerzas que ejercen elementos infinitesimales del fluido sobre otros elementos cercanos. Por ejemplo los elementos diagonales $T^{(i)(i)}$ (sin sumatorio, por esto ponemos los índices entre paréntesis) representan la componente F^i de la fuerza sobre la superficie $x^i = cte$, o sea la presión P^i en la dirección x^i.

Aclaramos estos conceptos generales un poco más con unos ejemplos concretos. El tensor de energía-momento de un conjunto de partículas que no interaccionan, también llamado *materia fría* o *materia no-relativista*, viene dado por

$$T^{\mu\nu}_{(\mathrm{mat})} = \rho_0 \, u^\mu u^\nu, \tag{10.1}$$

donde ρ_0 es la densidad de las partículas en el sistema de referencia en reposo con respecto a las partículas y u^μ es la cuadrivelocidad, definida en (5.27). En el sistema de referencia de un *observador comóvil*, que se mueve junto con el fluido, es decir, que tiene el mismo vector de velocidad u^μ que el fluido alrededor de él, la única componente non-nula es $T^{00}_{(\mathrm{mat})} \stackrel{*}{=} \rho_0$, es decir la densidad de materia. En otras palabras, la materia no-interactiva no tiene presión. La ley de conservación de energía (masa) dice que en estas coordenadas $\partial_t \rho_0 \stackrel{*}{=} 0$, ó en componentes del tensor de energía-momento, $\partial_0 T^{00}_{(\mathrm{mat})} \stackrel{*}{=} 0$. En cualquier otro sistema de referencia, relacionado a través de una transformación de Lorentz con el sistema en reposo, esta expresión toma la forma

$$\partial_\mu T^{\mu\nu}_{(\mathrm{mat})} = 0. \tag{10.2}$$

No es difícil ver que la componente 0 de esta expresión corresponde con la ley de conservación de energía en forma de una ecuación de continuidad

$$\partial_0 T^{00}_{(\mathrm{mat})} + \partial_i T^{i0}_{(\mathrm{mat})} \; = \; \partial_t \rho + \partial_i (\rho v^i) = 0, \tag{10.3}$$

donde $\rho = \gamma^2 \rho_0$ es la densidad medida por el observador en movimiento uniforme con respecto a las partículas y ρv^i es la densidad de momento. La densidad de masa en una región aumenta o

disminuye por lo tanto por el flujo de materia (momento) que entra o sale. Las componentes i de la ecuación (10.2) son la ley de conservación de momento, ya que se traduce a

$$\partial_0 T^{0i}_{(\text{mat})} + \partial_j T^{ji}_{(\text{mat})} \;=\; \partial_t(\rho v^i) + \partial_j(\rho v^j v^i), \tag{10.4}$$

lo que corresponde a la versión de la ecuación de Euler en la mecánica de fluidos, para el caso de un fluido sin presión ni fuerzas externas.

Las leyes de conservación de energía y momento, codificadas en esta expresión son tan fundamentales que se puede tomar esta última expresión como una condición necesaria para un tensor de energía-momento físciamente relevante. Veremos en la sección 10.2 que la condición $\nabla_\mu T^{\mu\nu} = 0$ juega un papel importante en la formulación de las ecuaciones del campo gravitacional.

Un ejemplo un poco más complicado es el caso del fluido perfecto, con densidad ρ y presión P. El tensor de energía-momento en este caso viene dado por

$$T^{\mu\nu}_{(\text{perf.fl.})} = (\rho_0 + P)u^\mu u^\nu - P\eta^{\mu\nu}. \tag{10.5}$$

Un observador comóvil, verá localmente un fluido isótropo y para este observador, el tensor de energía-momento tiene la forma sencilla

$$T^{\mu\nu}_{(\text{perf.fl.})} \stackrel{*}{=} \begin{pmatrix} \rho_0 & 0 & 0 & 0 \\ 0 & P & 0 & 0 \\ 0 & 0 & P & 0 \\ 0 & 0 & 0 & P \end{pmatrix} \tag{10.6}$$

Se puede demostrar que, imponiendo $\partial_\mu T^{\mu\nu}_{(\text{perf.fl.})} = 0$ en (10.5) para el caso de espacio plano, encontramos las leyes de conservación de energía y de momento en la versión relativista:

$$\begin{aligned} &\partial_\mu(\rho_0 u^\mu) + P\partial_\mu u^\mu = 0, \\ &(\rho_0 + P)u^\mu \partial_\mu u^\nu - \partial^\nu P + \partial^\mu P u_\mu u^\nu = 0. \end{aligned} \tag{10.7}$$

En general la presión y la densidad del fluido no son variables independientes y la relación entre las dos se llama la *ecuación de estado*, $P = P(\rho)$. Muchas veces se supone una relación lineal entre la presión y la densidad

$$P = w\rho, \tag{10.8}$$

donde la constante de proporcionalidad w es el *parámetro de la ecuación de estado* . El tensor de energía-momento (10.5) y la ecuación de estado (10.8) son de gran importancia en la cosmología, donde los distintos tipos de energía están representados por fluidos perfectos con distintos valores de w. Por ejemplo, la materia se comporta como materia fría y está caracterizado por $P = 0$, mientras la radiación (o materia relativista) satisface $P = \frac{1}{3}\rho$ y una constante cosmológica $P = -\rho$. Entraremos en más detalle en el Capítulo 13.

Finalmente, un tensor de energía-momento del campo electromagnético viene dado por

$$T^{\mu\nu}_{(\text{em})} \;=\; -F^{\mu\rho}F^\nu{}_\rho \;+\; \frac{1}{4}\eta^{\mu\nu}F_{\rho\lambda}F^{\rho\lambda}. \tag{10.9}$$

Efectivamente, en lenguaje tridimensional en función de los campos eléctricos $\vec{E}$ y magnéticos $\vec{B}$, los componentes se traducen a (ejerc.)

$$\begin{aligned} T^{00}_{(\text{em})} &= \frac{1}{2}(E^2 + B^2), \qquad T^{0i}_{(\text{em})} \;=\; (\vec{E}\times\vec{B})_i, \\ T^{ij}_{(\text{em})} &= E_i E_i + B_i B_j \;+\; \frac{1}{2}\delta^{ij}(E^2 + B^2). \end{aligned} \tag{10.10}$$

En $T^{00}_{(\mathrm{em})}$ reconocemos la expresión para la densidsad de energía E_{em} del campo electromagnético (1.39a), mientras $T^{0i}_{(\mathrm{em})}$ son las componentes del vector de Poynting (1.39b), $\vec{S} = \vec{E} \times \vec{B}$, que describe el flujo de energía y del momento del campo electromagnético. Finalmente, las componentes $T^{ij}_{(\mathrm{em})}$ se pueden identificar con el tensor de momento electromagnético (1.45), que describe el flujo de momento del campo electromagnético. En lenguaje covariante, la densidad de energía, el vector de Poynting y el tensor $T^{ij}_{(\mathrm{em})}$ son componentes distintas del mismo tensor $T^{\mu\nu}_{(\mathrm{em})}$.

Si calculamos la divergencia del tensor de energía-momento (10.9) utilizando las leyes de Maxwell, vemos que en presencia de cargas y corrientes la divergencia no es cero, sino (ejerc.)

$$\partial_\mu T^{\mu\nu}_{(\mathrm{em})} = j_\mu F^{\mu\nu}. \tag{10.11}$$

Dado que antes hemos interpretado $\partial_\mu T^{\mu\nu} = 0$ como la ley de conservación de energía y de momento, está claro que (10.11) implica que la energía y el momento del campo electromagnético no se conserva, por lo menos no en presencia de cargas y corrientes. Esto es lógico, puesto que el campo electromagnético interacciona con las partículas cargadas y es capaz de acelerarlas y cambiar su energía y su momento.

Efectivamente, escribiendo (10.11) en lenguaje tridimensional vemos que la componente temporal se traduce como

$$\begin{aligned} 0 &= \partial_0 T^{00}_{(\mathrm{em})} + \partial_i T^{i0}_{(\mathrm{em})} - j_i F^{0i} \\ &= \frac{1}{2}\partial_t (E^2 + B^2) + \vec{\nabla} \cdot (\vec{E} \times \vec{B}) + \vec{\jmath} \cdot \vec{E} \\ &= \partial_t \Big(E_{\mathrm{em}} + E_{\mathrm{cin}} \Big) + \vec{\nabla} \cdot \vec{S}, \end{aligned} \tag{10.12}$$

donde en la última igualdad hemos utilizado que el cambio en energía cinética E_{cin} de las partículas cargadas es justo el trabajo realizado por el campo eléctrico. A través de (3.26) tenemos que

$$\frac{dE_{\mathrm{cin}}}{dt} = \vec{v} \cdot \vec{F} = \sum_a q_a \vec{v} \cdot \vec{E} = \vec{\jmath} \cdot \vec{E}. \tag{10.13}$$

La ecuación (10.12) dice por lo tanto que el cambio de energía total en un volumen, siendo la suma de la energía del campo electromagnético y la energía cinética de las partículas cargadas, es igual al flujo de energía a través de la superficie. En otras palabras, en presencia de cargas y corrientes, la energía del campo electromagnético no se conserva, pero la energía total (la electromagnéctica y la de la materia) sí.

Del mismo modo podemos interpretar la componente espacial de (10.11) como la ley de conservación de momento total:

$$\begin{aligned} 0 &= \partial_0 T^{0i}_{(\mathrm{em})} + \partial_k T^{ki}_{(\mathrm{em})} - j_0 F^{0i} - j_k F^{ki} \\ &= \partial_t S_i + \partial_k T^{ik} + \rho E_i + \epsilon^{ikl} j_k B_l \\ &= \partial_t \Big(S_i + \sum_a (p_a)_i \Big) + \partial_k T^{ik}, \end{aligned} \tag{10.14}$$

donde en la última igualdad hemos escrito la fuerza de Lorentz (3.35) como el cambio de momento de las partículas cargadas, a través de la segunda ley de Newton (3.24). Por lo tanto, el cambio de momento del campo electromagnético más el cambio de momento de las partículas es igual al flujo de momento por la superficie, de modo que la cantidad total de momento se conserva.

Por lo tanto, la divergencia del tensor de energía-momento (10.9) del campo electromagnético no es cero, porque no describe un sistema cerrado, sino uno que interacciona con las partículas (o

los campos) cargadas, que también contribuyen al tensor de energía-momento total. El sistema total está descrita por el lagrangiano completo,

$$\mathcal{L} \ = \ \mathcal{L}_{(\text{em})} \ + \ \mathcal{L}_{(\text{carga})} \ + \ \mathcal{L}_{(\text{interacc})}, \tag{10.15}$$

que consiste de un término que describe la dinámica del campo electromagnético, uno que describe la dinámica de los campos cargados y un término de interacción. En la sección 10.3 veremos que cada uno de estos términos contribuye al tensor de energía-momento total,

$$T^{\mu\nu}_{(\text{tot})} \ = \ T^{\mu\nu}_{(\text{em})} \ + \ T^{\mu\nu}_{(\text{carga})} \ + \ T^{\mu\nu}_{(\text{interacc})}, \tag{10.16}$$

que está conservado en su totalidad, $\partial_\mu T^{\mu\nu}_{(\text{tot})} = 0$, pero las interacciones entre los distintos sectores de la teoría hacen que no se conservan los tensores de energía-momento de cada sector por separado.

10.2. Las ecuaciones de Einstein

Resumamos la situación hasta ahora: por un lado sabemos del Principio de Equivalencia que la gravedad es una manifestación de la curvatura del espacio, es decir, de una propiedad geométrica del espaciotiempo. Por otra lado sabemos que la fuente de esta curvartura es la materia de la cual tenemos una descripción tensorial, el tensor de energía-momento. Pero todavía no sabemos exactamente cómo la materia interacciona con el espaciotiempo. Esta interacción viene dada por las *ecuaciones de Einstein*.

La pregunta ahora claramente es: ¿Cuál es la forma exacta de las ecuaciones de Einstein? En otras palabras, ¿cómo podemos describir de manera cualitativa la interacción entre el espaciotiempo y la materia? El Principio de Covariancia nos dice que la ecuación debe ser válida en todos los sistemas de referencia, y que por lo tanto debe tener una forma tensorial. Concretamente, la ecuación de Einstein tiene que ser de la forma

$$G_{\mu\nu} = -\kappa T_{\mu\nu}, \tag{10.17}$$

donde $G_{\mu\nu}$ es un tensor que describe la curvatura del espacio, $T_{\mu\nu}$ el tensor de energía-momento y κ una constante de proporcionalidad (introducimos el signo menos para futura conveniencia).

La pregunta por lo tanto se reduce a la identificación del tensor $G_{\mu\nu}$. Resulta que hay diversas restricciones matemáticas y físicas que $G_{\mu\nu}$ tiene que cumplir:

1. $G_{\mu\nu}$ tiene que ser simétrico en los dos índices, ya que $T_{\mu\nu}$ también lo es.
2. $G_{\mu\nu}$ tiene que ser un objeto puramente geométrico. Por lo tanto, tiene que ser una función solamente de la métrica $g_{\mu\nu}$ y sus derivadas.
3. Para el espacio plano, tenemos que $G_{\mu\nu} = 0$.
4. La ley de conservación de energía $\nabla_\mu T^{\mu\nu} = 0$ implica que también $\nabla_\mu G^{\mu\nu} = 0$.
5. Se puede identificar la componente g_{00} de la métrica con el potencial gravitacional newtoniano (véase sección 11.1). Para tener una teoría dinámica y para recuperar la ecuación de Poisson (5.58), $G_{\mu\nu}$ debe contener segundas derivadas de la métrica. La manera más natural, por lo tanto es a través de las contracciones del tensor de Riemann $R_{\mu\nu\rho}{}^{\lambda}$.
6. Para obtener una ecuación diferencial de segundo orden (y no más) en los potenciales gravitatorios, $G_{\mu\nu}$ tiene que ser lineal en el tensor de Riemann. Contracciones del tipo $R_{\mu\rho}R_{\nu}{}^{\rho}$

ó $R_{\mu\rho\lambda\sigma}R_{\nu}{}^{\rho\lambda\sigma}$ darían lugar a ecuaciones diferenciales de orden más alto que 2.[1]

Posibles candidatos más o menos obvios para $G_{\mu\nu}$ podrían ser la misma métrica $g_{\mu\nu}$, su d'Alambertiano $\nabla_\rho\nabla^\rho g_{\mu\nu}$ o el tensor de Ricci $R_{\mu\nu}$, pero pensando un poco uno se da cuenta en seguida de que ninguna de estas posibilidades cumple todas las condiciones mencionadas arriba. Aunque la métrica tiene el rango y las simetrías adecuadas y satisface la condición $\nabla_\mu g^{\mu\nu} = 0$, tiene la desventaja de que no cumple la condición 5: la ecuación $g_{\mu\nu} = -\kappa T_{\mu\nu}$ no es una ecuación dinámica, ni mucho menos recupera la ecuación de Poisson (5.58). El d'Alambertiano $\nabla_\rho\nabla^\rho g_{\mu\nu}$ sufre del problema opuesto, ya que satisface (casi) todas las condiciones, pero es identicamente cero, por el hecho de que la conexión de Levi-Civita es compatible con la métrica, como contamos en la sección 8.1. Finalmente, $R_{\mu\nu}$ no satisface la condicón 4, sino $\nabla_\mu R^{\mu\nu} = \frac{1}{2}\partial^\nu R$, como vimos en (8.13). Por lo tanto, la ley de conservación de energía impondría que las únicas métricas permitidas serían las que tienen $\partial_\mu R = 0$, lo que no es una realista de esperar.[2]

En realidad las condiciones 1 - 6 determinan el tensor $G_{\mu\nu}$ unívocamente: se puede demostrar que la expresión más general para un tensor simétrico de rango $(0,2)$, construido de la métrica y sus derivadas y lineal en $R_{\mu\nu\rho\lambda}$ es, salvo una constante común, de la forma

$$G_{\mu\nu} = R_{\mu\nu} \;+\; \alpha g_{\mu\nu}R \;+\; g_{\mu\nu}\Lambda(x), \tag{10.18}$$

con α una constante y $\Lambda(x)$ una función escalar con dimensiones $\mathbf{ML}^{-3}$. Exigir que $\nabla_\mu G^{\mu\nu} = 0$ implica que $\alpha = -1/2$ y que Λ es una constante, mientras que exigir que $G_{\mu\nu} = 0$ para el espacio plano implica que $\Lambda = 0$. Por lo tanto el único tensor que satisface todas las condiciones necesarios es el tensor de Einstein, introducido en (7.40),

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R. \tag{10.19}$$

Una comparación con las fórmulas newtonianas (vease sección 11.1) fija la constante de proporcionalidad $\kappa = 8\pi G_N$, donde G_N es la constante de Newton, de modo que las ecuaciones de Einstein vienen dadas por

$$\boxed{R_{\mu\nu} \;-\; \frac{1}{2}g_{\mu\nu}R \;=\; -8\pi G_N T_{\mu\nu}.} \tag{10.20}$$

Las ecuaciones de Einstein forman un sistema de 10 ecuaciones diferenciales parciales nolineales acopladas de segundo orden, lo que hace que sean muy difíciles de resolver analíticamente. No hay técnicas conocidas para obtener una solución general. Todas las soluciones conocidas son casos con mucha simetría u obtenidas a través de técnicas específicas. Comentaremos más sobre este asunto en el Capítulo 12.

Las ecuaciones de Einstein tiene 10 componentes, pero en realidad la condición $\nabla_\mu G^{\mu\nu} = 0$ impone 4 ligaduras, de modo que sólo 6 ecuaciones son realmente independientes. Esto implica que de las 10 componentes de la métrica sólo 6 están determinadas por las ecuaciones de Einstein y corresponden a grados de libertad físicos. Las otras 4 componentes son componentes no-físicas que expresan la libertad de elección de sistema de coordenadas. Si $g_{\mu\nu}$ es una solución de (10.20)

[1] En 1938, el físico húngaro Cornelius Lanczos (1893 - 1977) demostró que la variación del Lagrangiano de Gauss-Bonnet,

$$\mathcal{L} = \sqrt{|g|}\Big(R^2 - 4R_{\mu\nu}R^{\mu\nu} + R_{\mu\nu\rho\lambda}R^{\mu\nu\rho\lambda}\Big),$$

también da lugar a ecuaciones diferenciales de segundo orden. Sin embargo en cuatro dimensiones el término de Gauss-Bonnet es un término topológico (es la característica de Euler cuadrimensional) y por lo tanto no contribuye a las ecuaciones de movimiento. La acción que consideró Lanczos sí contribuye a la dinámica en 5 o más dimensiones.

[2] En 1915, no mucho antes de dar con la versión correcta, Einstein propuso $R_{\mu\nu} = -\kappa T_{\mu\nu}$ para la ecuaciones adecuadas para la gravedad. Einstein mismo se autocriticó de manera irónica: *"Este tipo hace lo que le conviene. Cada año retira lo que escribió en año anterior."*

expresada en coordenadas x^μ, las misma métrica $g_{\alpha\beta}$ expresada en coordenadas y^α también debería ser una solución. Esto sólo es posible si la métrica contiene 4 grados de libertad que no están determinados por las ecuaciones de Einstein y que representan la libertad de aplicar un cambio de coordenadas $x^\mu \to y^\alpha = y^\alpha(x^\mu)$.[3]

Esto es muy similar a los grados de libertad del potencial A_μ en la teoría de Maxwell: aunque A_μ tiene 4 componentes, la relación de dispersión (3.32) y la libertad de elección de gauge $A'_\mu = A_\mu + \partial_\mu \Lambda$ imponen dos condiciones, de modo que el potencial sólo tiene 2 grados de libertad físicos. La libertad de elección de gauge en electromagnetismo y la libertdad de elección de sistema de coordenadas en relatividad general tienen muchas analogías (véase Capítulo **??**). Igual que en electromagnetismo se puede eleminar esta libertad fijando el gauge (por ejemplo el gauge de Lorenz $\partial_\mu A^\mu = 0$), en relatividad general se puede fijar el sistema de coordenadas imponiendo unas condiciones sobre la métrica (por ejemplo la condición $g_{0\mu} \stackrel{*}{=} \delta^0_\mu$ para coordenadas localmente inerciales). En ambos casos lo que uno está haciendo es imponer condiciones sobre los grados de libertad no-físicos.

La gran diferencia conceptual entre las ecuaciones de Einstein y la teoría newtoniana de la gravedad es que las ecuaciones de Einstein describen la gravedad como una teoría de campos. En el Capitulo 1 hemos visto que el concepto de campo físico fue introducido por Faraday y aprovechado por Maxwell en el contexto del electromagnetismo, para resolver el problema de la acción a distancia. Donde en la ley de Newton o de Coulomb las partículas tienen interacciones (gravitacionales o electromagnéticas) a distancia, en una teoría de campos las partículas interaccionan indirectamente, a través de un campo que se extiende por el espacio y que sirve de intermediario para la interacción entre las partículas. Una perturbación se transmite a través del campo a velocidad finita (la velocidad de la luz en el caso del electromagnetismo y la gravedad). De este modo la relatividad general resuelve el problema de acción inmediata y a distancia en la gravedad newtoniana, comentado en la sección 5.4.

Pero también hay una gran diferencia conceptual entre la teoría de Maxwell y la relatividad general. En la teoría de Maxwell el campo intermediario es el campo electromagnético $F_{\mu\nu}$ (o los potenciales A_μ), que viven en un espaciotiempo específico, generalmente Minkowski (aunque no necesariamente). Sin embargo en relatividad general, el campo intermediario es justamemente la métrica $g_{\mu\nu}$, el tensor que resume todas las propiedades geométricas del espaciotiempo. Esto no sólo implica que el espaciotiempo es algo dinámico, que interacciona con la materia y consigo mismo, sino también que la geometría no está fija y a priori determinada. En contraste con la teoría de Maxwell (o sus generalizaciones, como Yang-Mills o el Modelo Estándar), en relatividad general el espaciotiempo no es un escenario estático dentro del cual ocurre la física, sino que es una parte activa del juego. En cierto modo, en la relatividad especial el espacio de Minkowski era un espaciotiempo absoluto, en el sentido de que no se contempla la posibilidad de otra geometría, ni de que se vuelva dinámico. Aquí la geometría no está determinada a priori, sino por el contenido de energía y materia y por las condiciones iniciales. Donde la relatividad especial eliminó el fantasma del espacio y el tiempo absoluto, la relatividad general acabó con la idea del espaciotiempo estático y la geometría dada a priori.

A veces es útil rescribir las ecuaciones de Einstein sin la traza . Tomando la traza de (10.20), es decir contrayendo con $g^{\mu\nu}$, encontramos

$$R = \kappa T, \qquad (10.21)$$

donde $T = g^{\mu\nu} T_{\mu\nu}$ y hemos utilizado que $g^{\mu\nu} g_{\mu\nu} = 4$ en 4 dimensiones. Sustituyendo esto en

[3]Es justo debido a esta dificultad técnica que Einstein dudaba si covariancia general era compatible con causalidad. En particular le preocupaba que unas ecuaciones covariantes no determinan por completo la forma de la métrica a partir de unas condiciones iniciales. Hoy en día sabemos que esta indeterminación no es debida a una falta de determinismo, sino que es una consecuencia de la libertad inherente de hacer cambios generales de coordenadas: los grados de libertad físicas de la métrica si están determinadas por las ecuaciones de Einstein y las condiciones iniciales, pero éstas nunca pueden determinar qué sistema de coordenadas vamos a usar para describir la solución.

(10.20) vemos que las ecuaciones de Einstein sin traza son de la forma

$$R_{\mu\nu} = -\kappa\Big(T_{\mu\nu} - \frac{1}{2}g_{\mu\nu}T\Big). \tag{10.22}$$

Esta ecuación es completamente equivalente a (10.20), pero es un poco más fácil a la hora de buscar soluciones, ya que no hace falta calcular el escalar de Ricci R. Históricamente, esta es la forma original en que Einstein escribió las ecuaciones, aunque su forma más famosa es sin duda (10.20).

Una de las ventajas de (10.22) es que en el vacío, donde $T_{\mu\nu} = 0$, las ecuaciones se reducen a

$$R_{\mu\nu} = 0. \tag{10.23}$$

Obviamente el espacio de Minkowski es una solución de esta ecuación (la condión 3 sobre $G_{\mu\nu}$), pero (10.23) también es suficientemente complicado para admitir soluciones no-triviales, como la solución de Schwarzschild o de ondas gravitacionales. Las soluciones de (10.23) son en cierto modo el análogo de las ondas electromagnéticas en teoría de Maxwell, que también son soluciones de las ecuaciones en el vacío. Las métricas que tienen la propiedad (10.23) se llaman *Ricci-planas*.

Finalmente, antes hemos visto que la forma más general de un tensor simétrico de rango 2, linear en $R_{\mu\nu\rho\lambda}$ es de la forma (10.18) y que la condición de que $G_{\mu\nu}$ fuera cero para el espacio plano fijaba el valor de la constante Λ en cero. Si aflojamos esta última condición y no exigimos encontrar el espacio de Minkowski entre las posible soluciones, vemos que aparece un parámetro nuevo en las ecuaciones: la *constante cosmológica* Λ. La constante cosmológica no apareció en la primera publicación de Einstein, sino que la introdujo cuando empezaba a interesarse por la cosmología, con el fin de obtener un universo estático. El valor de la constante cosmológica no está determinado por la teoría y representa una fuerza universal repulsiva si $\Lambda > 0$ y atractiva si $\Lambda < 0$. Se la puede interpretar como la energía del vacío y muchas veces se la considera parte del tensor de energía-momento, más que de la parte geométrica de la ecuación de Einstein. Efectivamente el tensor de energía-momento correspondiente sería

$$T^{(\Lambda)}_{\mu\nu} = \Lambda g_{\mu\nu}, \tag{10.24}$$

en otras palabras un fluido perfecto con $\Lambda = \rho = -P$. Observaciones cosmológicas recientes parecen sugerir que en nuestro universo la constante cosmológica tiene un pequeño valor positivo y forma un 70 % del contenido de energía del universo (véase Capítulo **??**).

10.3. Física en espacios curvos y la acción de Einstein-Hilbert

Hemos visto que las ecuaciones de Einstein (10.20) nos dicen cómo la materia y la energía determinan la curvatura del espaciotiempo, pero no dicen nada sobre la dinámica de la materia y los campos no-gravitacionales. Obsérvese que (10.20) es una ecuación diferencial para $g_{\mu\nu}$, donde $T_{\mu\nu}$ aparece como un término inhomogeneo, no como algo dinámico. Lo que nos gustaría tener es una generalización a espacios curvos de la dinámica relativista de la sección 3.3, es decir de la Segunda Ley de Newton y de la teoría de Maxwell.

El Principio de Covariancia dice que las leyes de la física en un espacio curvo tienen que transformar bien bajo cambios generales de coordenadas, es decir tienen que estar escritas en términos de derivadas covariantes, en lugar de derivadas parciales. Y el Principio de Equivalencia dice que las leyes tienen que ser de tal forma que en las coordenadas localmente inerciales tienen que recuperar la forma de la relatividad especial. Pero aún así hay muchas maneras de generalizarlas.

Por ejemplo cualquiera de las expresiones

$$\begin{aligned}
&\nabla_\mu F^{\mu\nu} = j^\nu, \\
&\nabla_\mu F^{\mu\nu} + \nabla_\mu R_\rho{}^\nu F^{\mu\rho} = j^\nu, \\
&\nabla_\mu F^{\mu\nu} + R_{\mu\rho\lambda}{}^\nu \nabla^\lambda F^{\mu\rho} = j^\nu + R^{\mu\nu} j_\mu, \\
&\ldots
\end{aligned} \tag{10.25}$$

es una expresión covariante que se reduce en el espacio de Minkowski a la ley de Maxwell $\partial_\mu F^{\mu\nu} = j^\nu$. Sin embargo, físicamente no son equivalentes, por el hecho de que tienen acoplos distintos entre los campos electromagnéticos y la métrica y la curvatura.

¿Cómo sabemos cuál de estas generalizaciones es la correcta? Verificación experimental, por supuesto, ya que las distintas generalizaciones no son equivalentes. Pero comprobar experimentalmente la diferencia entre estas ecuaciones es muy difícil, puesto que para campos gravitatorios no muy fuertes, los términos proporcionales a los tensores de curvatura son muy pequeños. Estaría bien si hubiera algún principio físico que nos diera una prescripción para generalizar las fórmulas de relatividad especial.

El *Principio de Mínimo Acoplo* es lo más parecido a lo que uno puede aspirar de tal prescripción. Estrictamente hablando no es un principio físico, sino más bien un principio filosófico, una variante de la Navaja de Ockham. Básicamente, el Principio de Mínimo Acoplo supone que la generalización más sencilla es la correcta.

Principio del Mínimo Acoplo (formulación física): *En un espacio curvo, los campos no-gravitacionales se acoplan sólamente a la métrica, no al tensor de Riemann o sus contracciones.*

Por lo tanto, la prescripción para generalizar las leyes de la físcia a espacio curvos es muy simple y concreta:

Principio del Mínimo Acoplo (formulación práctica): *Las leyes de la física en espacios curvos son los mismos que las de relatividad especial, donde se sustituyen las derivadas parciales ∂_μ por covariantes ∇_μ y la métrica de Minkowski $\eta_{\mu\nu}$ por una métrica general $g_{\mu\nu}$.*

Como ya dijimos, no hay una razón física fundamental que justifique este principio, pero es atractivo por su sencillez. Sin embargo, merece la pena enfatizar que sólo es una idea elegante y que hay que aplicarla con mucha cautela. En particular la prescripción no siempre es unívoca, como por ejemplo en el caso del operador $\partial_\mu \partial_\nu$. En el espacio de Minkowski no importa el orden de las derivadas, pero al sustituirlas por derivadas covariantes si hay diferencia entre $\nabla_\mu \nabla_\nu$ y $\nabla_\nu \nabla_\mu$, debido a (8.1).

Aún así, el Principio de Mínimo Acoplo es suficientemente fuerte como para poder aplicarlo en la mayoría de los casos. Por ejemplo la segunda ley de Newton para una partícula libre $\ddot{x}^\mu = 0$ se convierte en

$$\frac{\nabla}{d\tau}\Big(\frac{dx^\mu}{d\tau}\Big) \;\equiv\; \ddot{x}^\mu + \Gamma^\mu_{\nu\rho} \dot{x}^\nu \dot{x}^\rho \;=\; 0, \tag{10.26}$$

lo que es justamente la ecuación (7.42) para una geodésica. En otras palabras, en un espacio curvo, una partícula sobre la cual no actúa ninguna fuerza aparte de la gravedad (es decir, una partícula libre) se mueve a lo largo de las geodésicas del espacio. Si la partícula si está sometida a fuerzas externas f^μ (no-gravitacionales), la segunda ley de Newton tiene la forma

$$m_0\Big(\ddot{x}^\mu + \Gamma^\mu_{\nu\rho} \dot{x}^\nu \dot{x}^\rho\Big) \;=\; f^\mu. \tag{10.27}$$

De la misma manera podemos generalizar la teoría de Maxwell a espacios curvos. Las leyes de Maxwell vienen dadas por

$$\nabla_\mu F^{\mu\nu} = j^\nu, \qquad\qquad \nabla_\mu F_{\nu\rho} + \nabla_\rho F_{\mu\nu} + \nabla_\nu F_{\rho\mu} = 0, \tag{10.28}$$

aunque la simetría de la conexión de Levi-Civita y la antisimetría de $F_{\mu\nu}$ permiten rescribir la ecuación homogénea como (ejerc.)

$$\partial_\mu F_{\nu\rho} + \partial_\rho F_{\mu\nu} + \partial_\nu F_{\rho\mu} = 0, \tag{10.29}$$

donde $|g|$ es el valor absoluto del determinante de la métrica.

La ecuación homogénea (10.29) tiene la misma forma que en el espacio de Minkowski (5.49) y sugiere por lo tanto que el tensor electromagnético sigue teniendo la misma forma con derivadas parciales, $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. Efectivamente debido a la simetría de la conexión y la antisimetría del tensor electromagnético tenemos que

$$\begin{aligned} F_{\mu\nu} &= \nabla_\mu A_\nu - \nabla_\nu A_\mu \\ &= \partial_\mu A_\nu - \Gamma^\rho_{\mu\nu} A_\rho - \partial_\nu A_\mu + \Gamma^\rho_{\nu\mu} A_\rho \\ &= \partial_\mu A_\nu - \partial_\nu A_\mu. \end{aligned} \tag{10.30}$$

De la mecánica analítica sabemos las ventajas del formalismo lagrangiano: variando con respecto a los grados de libertad se sacan las ecuaciones de movimiento, de modo que el lagrangiano en cierta forma es un resumen de la dinámica del sistema o de la teoría. Además muchas veces es más sencillo encontrar las simetrías del sistema en el lagrangiano que en las ecuaciones. Sería por lo tanto interesante encontrar un lagrangiano cuyas ecuaciones de movimiento fuesen justo las ecuaciones Newton, de Einstein y de Maxwell.

No es muy difícil encontrar una acción que nos proporcione la segunda ley de Newton (10.27), por lo menos en el caso donde las fuerzas externas son conservativas. Utilizando la expresión (8.7) para la relación entre la conexión de Levi-Civita y la métrica, es fácil ver que la ecuación de Euler-Lagrange de x^μ de la acción

$$S = \int d\tau \Big[-\tfrac{1}{2} m_0 g_{\mu\nu} \dot{x}^\mu \dot{x}^\nu - V(x_\mu x^\mu) \Big] \tag{10.31}$$

da justamente la segunda ley de Newton (10.27) con $f_\mu = -\partial_\mu V$ (ejerc.).

Una acción para las ecuaciónes de Einstein es un poco más difícil de encontrar. Concentrémosnos primero en el caso más sencillo de las ecuaciones del vacío, donde el tensor de energía-momento es cero. Una acción para este caso no fue encontrado por Einstein, sino por Hilbert y enviado junto con las ecuaciones de Einstein a una revista, 5 días antes de que el propio Einstein enviara el artículo en el que por fin daba con la versión correcta de sus ecuaciones. La *acción de Einstein-Hilbert*, que genera las ecuaciones de Einstein para el vacío, viene dada por

$$S = \frac{1}{2\kappa} \int d^4x \, \sqrt{|g|} \, R, \tag{10.32}$$

donde R es el escalar de Ricci, definido en (7.39).

En principio se obtienen las ecuaciones de Einstein (10.20) a través de la variación de la acción (10.32) con respecto a la métrica inversa $g^{\mu\nu}$. Sin embargo, en la práctica este método es muy complicado. Nótese que el escalar de Ricci es cuadrático y de primer orden en los símbolos de Christoffel,

$$R = g^{\mu\nu} \Big(\partial_\mu \Gamma^\lambda_{\lambda\nu} - \partial_\lambda \Gamma^\lambda_{\mu\nu} + \Gamma^\lambda_{\mu\sigma} \Gamma^\sigma_{\lambda\nu} - \Gamma^\lambda_{\mu\nu} \Gamma^\sigma_{\lambda\sigma} \Big), \tag{10.33}$$

mientras los símbolos de Christoffel a su vez son cuadráticos y de primer orden en la métrica

$$\Gamma^{\rho}_{\mu\nu} = \frac{1}{2} g^{\rho\lambda}\Big(\partial_\mu g_{\lambda\nu} + \partial_\nu g_{\mu\lambda} - \partial_\lambda g_{\mu\nu}\Big), \tag{10.34}$$

de modo que (10.32) es de tercer orden en $g^{\mu\nu}$. La variación de S lleva por lo tanto a una ecuación de Euler-Lagrange modificada, de cuarto orden:

$$\partial_\rho \partial_\lambda \Big(\frac{\delta \mathcal{L}}{\delta(\partial_\rho \partial_\lambda g^{\mu\nu})}\Big) \; - \; \partial_\rho \Big(\frac{\delta \mathcal{L}}{\delta(\partial_\rho g^{\mu\nu})}\Big) \; + \; \frac{\delta \mathcal{L}}{\delta g^{\mu\nu}} = 0, \tag{10.35}$$

donde $\mathcal{L}$ es el lagrangiano de la acción de Einstein-Hilbert (10.32). Estará claro que calcular explícitamente todos las componentes de (10.35), aunque posible, no es en la práctica un método viable.

Afortunadamente existe un truco, llamado el *formalismo de Palatini*, que nos permite obtener las ecuaciones de Einstein de manera más directa. El truco consiste en suponer por un momento que la conexión en el escalar de Ricci (10.33) es una conexión arbitraria, no necesariamente la conexión de Levi-Civita (10.34). En este caso no hay una relación entre la conexión y la métrica y se considera por lo tanto los $\Gamma^{\rho}_{\mu\nu}$ como un campo independiente. La acción de Einstein-Hilbert tiene por lo tanto sólo dos sitios donde aparece la métrica: en el determinante $\sqrt{|g|}$ y en la contracción del tensor de Ricci $R(\Gamma) = g^{\mu\nu} R_{\mu\nu}(\Gamma)$. La variación con respecto a $g^{\mu\nu}$ es entonces muy sencillo

$$0 \equiv \frac{\delta \mathcal{L}}{\delta g^{\mu\nu}} \;=\; \sqrt{|g|}\Big[R_{\mu\nu}(\Gamma) - \frac{1}{2} g_{\mu\nu} g^{\rho\lambda} R_{\rho\lambda}(\Gamma)\Big], \tag{10.36}$$

donde el primer término viene de la variación de R y el segundo de la variación del determinante. Dividiendo esta igualdad por $\sqrt{|g|}$ se tiene directamente las ecuaciones de Einstein (10.20) para el caso de $T_{\mu\nu} = 0$. Sin embargo, hay otros grados de libertad más, dado que hemos considerado la conexión como un campo independiente. Tenemos por lo tanto que calcular todavía la variación de la acción δS bajo variaciones de la conexión $\Gamma^{\rho}_{\mu\nu} \rightarrow \Gamma^{\rho}_{\mu\nu} + \delta\Gamma^{\rho}_{\mu\nu}$. No es difícil de averiguar que la variación de la acción viene dada por

$$\delta S = \frac{1}{2\kappa} \int d^4x \; \sqrt{|g|} \; g^{\mu\nu} \Big[\nabla_\mu(\delta\Gamma^{\lambda}_{\lambda\nu}) - \nabla_\lambda(\delta\Gamma^{\lambda}_{\mu\nu}) + T^{\sigma}_{\mu\lambda}(\delta\Gamma^{\lambda}_{\sigma\nu})\Big] \tag{10.37}$$

gracias a la *Identidad de Palatini*

$$\begin{aligned} \delta R_{\mu\nu} &= \partial_\mu(\delta\Gamma^{\lambda}_{\lambda\nu}) - \partial_\lambda(\delta\Gamma^{\lambda}_{\mu\nu}) + (\delta\Gamma^{\lambda}_{\mu\sigma})\Gamma^{\sigma}_{\lambda\nu} + \Gamma^{\lambda}_{\mu\sigma}(\delta\Gamma^{\sigma}_{\lambda\nu}) - (\delta\Gamma^{\lambda}_{\mu\nu})\Gamma^{\sigma}_{\lambda\sigma} - \Gamma^{\lambda}_{\mu\nu}(\delta\Gamma^{\sigma}_{\lambda\sigma}) \\ &= \nabla_\mu(\delta\Gamma^{\lambda}_{\lambda\nu}) - \nabla_\lambda(\delta\Gamma^{\lambda}_{\mu\nu}) + T^{\sigma}_{\mu\lambda}(\delta\Gamma^{\lambda}_{\sigma\nu}), \end{aligned} \tag{10.38}$$

donde $T^{\sigma}_{\mu\lambda} = \Gamma^{\sigma}_{\mu\lambda} - \Gamma^{\sigma}_{\lambda\mu}$ es el tensor de torsión (7.24) (Nótese que de momento tratamos la conexión como arbitraria, y por lo tanto no necesariamente simétrica). Integrando por partes los primeros dos términos, podemos reescribir (10.37) como (ejerc.)

$$\begin{aligned} \delta S \;=\; \frac{1}{2\kappa} \int d^4x \; \sqrt{|g|} \; \Big\{ & g^{\mu\nu} T^{\sigma}_{\mu\lambda}(\delta\Gamma^{\lambda}_{\sigma\nu}) \; - \; \Big[\nabla_\mu g^{\mu\nu} \; + \; \frac{1}{2} g^{\mu\nu} g^{\sigma\tau} \nabla_\mu g_{\sigma\tau} \; + \; g^{\mu\nu} T^{\sigma}_{\mu\sigma}\Big](\delta\Gamma^{\lambda}_{\lambda\nu}) \\ & + \Big[\nabla_\lambda g^{\mu\nu} \; + \; \frac{1}{2} g^{\mu\nu} g^{\sigma\tau} \nabla_\lambda g_{\sigma\tau} \; + \; g^{\mu\nu} T^{\sigma}_{\lambda\sigma}\Big](\delta\Gamma^{\lambda}_{\mu\nu})\Big\}, \end{aligned} \tag{10.39}$$

donde hemos utilizado la generalización a conexiónes arbitrarias de (A.22) (ejerc.),

$$\partial_\rho \sqrt{|g|} \;=\; \frac{1}{2}\sqrt{|g|} g^{\mu\nu} \partial_\rho g_{\mu\nu} \;=\; \frac{1}{2}\sqrt{|g|} g^{\mu\nu} \nabla_\rho g_{\mu\nu} + \sqrt{|g|}\Gamma^{\lambda}_{\rho\lambda}. \tag{10.40}$$

Por lo tanto, la variación de la acción se puede escribir como

$$\begin{aligned} 0 \;\equiv\; \delta S \;=\; \frac{1}{2\kappa} \int d^4x \; \sqrt{|g|} \; (\delta\Gamma^{\lambda}_{\mu\nu}) \Big\{ & -\Big[\nabla_\rho g^{\rho\nu} + g^{\rho\nu} g^{\sigma\tau} \nabla_\rho g_{\sigma\tau} + g^{\rho\nu} T^{\sigma}_{\rho\sigma}\Big]\delta^{\mu}_{\lambda} \\ & + \nabla_\lambda g^{\mu\nu} + g^{\mu\nu} g^{\sigma\tau} \nabla_\lambda g_{\sigma\tau} + g^{\mu\nu} T^{\sigma}_{\lambda\sigma} + g^{\rho\nu} T^{\mu}_{\rho\lambda}\Big\}. \end{aligned} \tag{10.41}$$

El hecho de que la variación δS tenga que ser cero para cualquier variación $\delta\Gamma^{\lambda}_{\mu\nu}$, implica que el integrando entre los corchetes es cero y dado que los términos dentro de los corchetes con proporcionales a la derivada covariante de la métrica y la torsión, una solución obvia es

$$T^{\mu}_{\rho\lambda} = 0, \qquad\qquad \nabla_\rho g_{\mu\nu} = 0. \tag{10.42}$$

En otras palabras, la acción es estacionaria bajo variaciones de la conexión, si la torsión y la derivada covariante de la métrica son cero. Pero la simetría la conexión y su compatibilidad con la métrica son justamente las dos condiciones de la sección 8.1 que determinan unívocamente la conexión de Levi-Civita. emos por lo tanto que la conexión de Levi-Civita surge de manera natural en la relatividad general como consecuencia de las ecuaciones de movimiento de una conexión general.

Resumiendo podemos decir que el formalismo de Palatini trata la métrica y la conexión como a priori independientes. La ecuación de movimiento de la métrica nos da las ecuaciones de Einstein, mientras la ecuación de la conexión identifica la conexión como la de Levi-Civita. Hasta ahora sólo hemos aplicado el formalismo de Palatini a la acción de Einstein-Hilbert, para obtener las ecuaciones del vacío, pero el formalismo se aplica de igual manera a acciones con acoplos a campos no-gravitacionales. La variación de la métrica dará las ecuaciones Einstein acoplado al tensor de energía-momento y la variación de la conexión dará las mismas condiciones, ya que los campos no-gravitacionales se acoplan a la métrica y no a la conexión. En la práctica por lo tanto podemos olvidarnos de esta última variación y suponer directamente que la conexión es la de Levi-Civita, variando sólo con respecto a $g_{\mu\nu}$.

Finalemente, para obtener también las ecuaciones de Maxwell tenemos que generalizar la acción de Einstein-Hilbert a la llamada *acción de Einstein-Maxwell*

$$S = \int d^4x \sqrt{|g|}\, \Big[\frac{1}{2\kappa} R \; - \; \frac{1}{4} F_{\mu\nu} F^{\mu\nu}\Big]. \tag{10.43}$$

Obsérvese que las contracciones de la parte electromagnética están hechas a través de la métrica general $g_{\mu\nu}$, de acuerdo con el Principio de Mínimo Acoplo, $F_{\mu\nu}F^{\mu\nu} = g^{\mu\rho}g^{\nu\lambda}F_{\mu\nu}F_{\rho\lambda}$.

La variación con respecto a $g^{\mu\nu}$ da la ecuación de Einstein en presencia de campos electromagnéticos (ejerc.)

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R \; = \; \kappa \Big[F_{\mu\rho} F_{\nu}{}^{\rho} - \frac{1}{4} g_{\mu\nu} F_{\rho\lambda} F^{\rho\lambda} \Big], \tag{10.44}$$

mientras la ecuación de Euler-Lagrange para A_μ,

$$\partial_\mu \Big(\frac{\delta \mathcal{L}}{\delta(\partial_\mu A_\nu)} \Big) \; - \; \frac{\delta \mathcal{L}}{\delta A_\nu} = 0, \tag{10.45}$$

nos proporciona la siguiente expresión para la ecuación de movimiento el campo electromagnético (ejerc.)

$$\frac{1}{\sqrt{|g|}} \partial_\mu \Big[\sqrt{|g|}\; F^{\mu\nu} \Big] = 0, \tag{10.46}$$

lo que, usando la simetría de la conexión, la antisimetría de $F^{\mu\nu}$ y la propiedad (8.30), podemos reescribir como la ecuación de Maxwell en espacio curvos en ausencia de fuentes, en forma explicitamente covaraiante (ejerc.),

$$\nabla_\mu F^{\mu\nu} = 0. \tag{10.47}$$

Nótese que el lado derecho de la ecuación de Einstein es justo (salvo un signo) el tensor de energía-momento (10.9). Esto es una propiedad general: Dado un lagrangiano $\mathcal{L} = \mathcal{L}_{\rm EH} + \mathcal{L}_\phi$, que consiste de la acción de Einstein-Hilbert (10.32) más los términos $\mathcal{L}_\phi$ que describen la dinámica de un campo ϕ que interacciona con la gravedad, entonces el tensor de energía-momento de ϕ se puede obtener a través de una variación de $\mathcal{L}_\phi$ con respecto a $g^{\mu\nu}$:

$$T_{\mu\nu} = \frac{2}{\sqrt{|g|}} \frac{\delta(\sqrt{|g|}\, \mathcal{L}_\phi)}{\delta g^{\mu\nu}}. \tag{10.48}$$

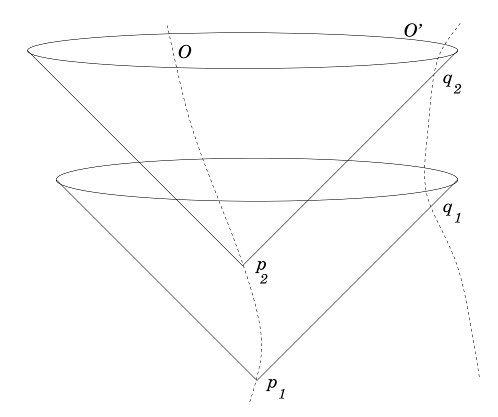


Figura 10.1: *La estructura causal del espacio de Minkowski: Dos observadores $\mathcal{O}$ y $\mathcal{O}'$ están continuamente en contacto causal. Un evento p en la historia de $\mathcal{O}$ influenciará a $\mathcal{O}'$ a partir del momento q en que $\mathcal{O}'$ entre en el cono de luz de p. Equivalentemente (pero no dibujado), los eventos en la historia de $\mathcal{O}'$ influenciarán a $\mathcal{O}$ desde el momento en que éste entra en su cono de luz.*

Obsérvese que la acción de Einstein-Maxwell (10.43) es invariante tanto bajo cambios generales de coordenadas como bajo transformaciones gauge.

10.4. La estructura local y global del espaciotiempo

Hemos visto que las ecuaciones de Einstein (10.20) relacionan la curvatura del espacio-tiempo con el contenido de energía y materia dentro del espacio y que la relación entre ambas cosas es tal que a escalas pequeñas se recuperan las leyes de la relatividad especial.

Esto implica que, por lo menos localmente, la estructura del espaciotiempo general es similar a la estructura de Minkowski. Por ejemplo, aunque en general la métrica tendrá una forma más complicada, la signatura siempre será $(+, -, -, -)$, es decir, el espaciotiempo siempre es lorentziano. Por lo tanto, igual que en el espacio de Minkowski, podemos, en cualquier punto p, dividir los vectores V^μ de $T_p(\mathcal{M})$ en tres clases, los vectores temporales, nulos y espaciales, dependiendo de si la norma $V^2 = g_{\mu\nu}V^{\mu\nu}V^\nu$ es positiva, cero o negativa. Gracias al hecho de que V^2 es un escalar, tendrá en mismo valor en todos los sistemas de coordenadas, de modo que la clasificación de los vectores es la misma para todos los observadores.

De manera análoga al espacio de Minkowski, en cualquier punto p se puede definir el cono de luz en un punto p como el conjunto de vectores nulos $V^2(p) = 0$ y este cono de luz determina la estructura causal del espaciotiempo. Pero la gran diferencia con el espacio de Minkowski es que, debido a la curvatura del espaciotiempo, la orientación del cono de luz no es la misma en todo el espaciotiempo, sino que varía de punto en punto. La interpretación física de este fenómeno es que la gravedad afecta la trayectoria de la luz, mientras matemáticamente hablando es debido al hecho de que las coordenadas localmente inerciales son distintas en cada punto de la variedad.

Una partícula masiva seguirá una trayectoria temporal, que se define como una curva cuyos vectores tangentes son temporales en todos los puntos de la curva. De la misma manera una partícula sin masa seguirá una curva nula. Dos eventos p y q están en contacto causal si están conectados a través de una curva temporal o nula. Por otro lado, si entre p y q no existe una curva temporal o nula, no es posible intercambiar información entre los dos eventos y se dice que los eventos están fuera de contacto causal.

En el espacio de Minkowski, un evento p en la historia de un observador $\mathcal{O}$ puede llegar

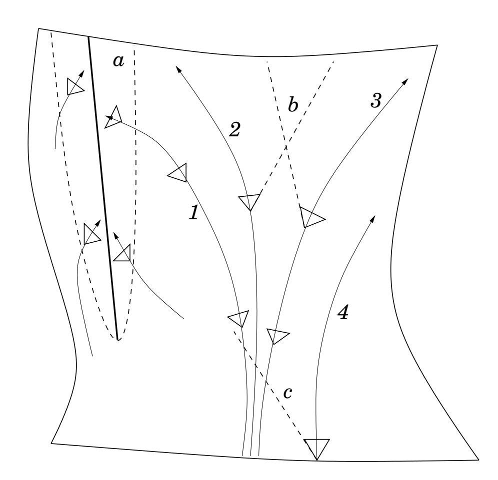


Figura 10.2: *Horizontes en espacios generales: Las partículas 1, 2 y 3 están inicialmente en contacto causal, pero a medida que pasa el tiempo ya no es posible mandar señales de luz entre los observadores. La partícula 1, al pasar un horizonte de sucesos a, entra en una región donde la luz se queda atrapada, debido a un fuerte campo gravitatorio. Por otro lado, las partículas 2 y 3 pierden contacto debido a la rápida aceleración del universo: señales emitidos tarde en la historia no llegan nunca de un observador al otro y cada observador tiene un horizonte cósmico b detrás del cual no puede ver. Las partículas 1, 2 y 3 no reciben influencias de 4 hasta que 4 aparece dentro del horizonte de partículas c, debido a la edad finita del universo.*

a influenciar tarde o temprano los hechos en cualquier posición (x, y, z) del espacio, ya que el cono de luz crece con el paso del tiempo. Cualquier observador $\mathcal{O}'$ (siempre y cuando no tenga aceleraciones demasiado grandes con respecto a $\mathcal{O}$) está continuamente influenciado por $\mathcal{O}$, ya que tarde o temprano entrará en el cono de luz de cualquier evento p en la linea de universo de $\mathcal{O}$ (véase Figura 10.1). Al revés también es verdad, tarde o temprano $\mathcal{O}$ entrará en el como de luz de cualquier evento en la historia de $\mathcal{O}'$. Se dice que estas observadores estan en *contacto causal*.

Sin embargo en un espaciotiempo general esto no necesariamente tiene que ser cierto. Debido al hecho de que la orientación de los conos de luz varía de punto o punto, es posible que unas partículas, que inicialmente estaban en contacto causal, dejen de estarlo después de un cierto tiempo (véase Figura 10.2). Si un observador $\mathcal{O}$ entra en una región del espacio desde la cual ya no es capaz de mandar señales de luz a un observador $\mathcal{O}'$ fuera de esta región, decimos que el espaciotiempo tiene un *horizonte*. Existen varios tipos de horizones y sólo comentaremos los más comunes.

Un tipo de horizonte surge cuando el campo gravitatorio es tan fuerte que la luz se queda atrapada. En términos de la física clásica (newtoniana), la velocidad de escape de esta región es más grande que la velocidad de la luz, así que nada que entre en esta región puede volver a escapar. Este tipo de horizonte se llama un *horizonte de sucesos* y aparece en la formación de agujeros negros (véase el caso a en Figura 10.2). Un horizonte de sucesos es un horizonte absoluto, en el sentido de que es el mismo para todos los observadores.

Otro tipo de horizonte no es absoluto, sino que depende del observador. En un universo en expansión, la expansión puede ser tan rápida que dos observadores "se pierde de vista", es decir la distancia entre los dos observadores aumenta demasiado rápido para que el la luz emitida por un observador llegue al otro.[4] Para cada observador hay un momento en su historia después del cual ya no es capaz de recibir señales del otro y decimos que hay un *horizonte cósmico* (caso b en

[4]Ingenuamente uno podría pensar que esto implica que la velocidad relativa entre los observadores es mayor que la velocidad de la luz, pero recordemos que en una variedad general, la velocidad relativa entre dos observadores distantes no está bien definida. El vector velocidad sólo está definido en el espacio tangente, es decir, en la práctica en una región pequeña alrededor de la partícula. Las velocidades relativas entre partículas sólo están definidas para partículas cercanas.

la Figura 10.2).

Finalmente, si el universo tiene una edad finita, ningún observador puede ver más alla de la distancia que ha recordido la luz desde el origen del universo hasta el momento de observación. Desde distancias más alla, la luz simplemente no ha tenido tiempo todavía para llegar hasta el observador, aunque sí llegará con el paso del tiempo. El horizonte asociado con este fenómeno se llama un *horizonte de partículas*. (caso c en la Figura 10.2). Horizontes cósmicos y horizontes de partículas son muy comunes en cosmología.

Muchas veces la métrica que surge como solución de las ecuaciones de Einstein es singular en cierta región del espaciotiempo, es decir, algunos componentes y/o el determinante de $g_{\mu\nu}$ tienden a cero ó a infinito. En este caso tenemos que distinguir entre dos tipos de singularidades. El primer tipo son las *singularidades de coordenadas*, que son simplemente artefactos del sistema de coordenadas que utilizamos. El origen en coordenadas esféricas en $\mathbb{R}^3$ es un ejemplo de una singularidad de coordenadas, puesto que el determinante $|g| = r^2 \sin^2\theta$ tiende a cero por $r \rightarrow 0$, aunque en coordenadas cartesianas está claro que el origen es un punto completamente regular.

Una *singularidad física* sin embargo es una singularidad real, un punto donde la curvatura del espacio es infinita. Cerca de una singularidad física, el campo gravitatorio es muy fuerte y los efectos de marea son muy grandes. Los físicos británicos Stephen Hawking (1942 -) y Robert Penrose (1931 -) demostraron en los años 60 que en el interior de una superficie que atrapa la luz, necesariamente tiene que haber una singularidad física, de modo que la singularidad en el centro de agujeros negros es física. Por otra parte, hay una conjectura, llamada el *censor cósmico*, de que cualquier singularidad física siempre está rodeada por un horizonte de sucesos, de modo que la singularidad misma no es visible desde fuera y no puede influenciar al resto del universo y crear problemas de causalidad. Singularidades que no están rodeados por horizontes se llaman *singularidades desnudas*.

La manera general para distinguir entre singularidades de coordenadas y singularidades físicas es calculando los *invariantes de curvatura*, unas funciones escalares formadas de contracciones del tensor de Riemann, como R, $R_{\mu\nu}R^{\mu\nu}$, $R_{\mu\nu\rho\lambda}R^{\mu\nu\rho\lambda}$, etc. Si los invariantes de curvatura son singulares, entonces la singularidad es una singularidad fisica. Por otro lado si los invariantes son regulares, es probable, pero no seguro que la singularidad sea de coordenadas. Nótese que siempre es posible que un invariante todavía no calculado se haga singular. La prueba definitiva de que una singularidad sea de coordenadas es encontrar un cambio de coordenadas que hace desaparecer la singularidad.

Las ecuaciones de Einstein son ecuaciones diferenciales y tienen por lo tanto un carácter local. La curvatura en cierto punto del espaciotiempo viene dada por el contenido de energía-momento en este punto y una pequeña región alrededor, pero la geometría global puede llegar a tener estructuras extrañas, contraintuitivas e incluso patológicas. Discutiremos brevemente unos cuantos ejemplos.

Una primera propiedad geométrica de un espaciotiempo es si el espacio es *geodésicamente completo*. Un espacio es geodésicamente completo si se puede extender el parametro afín de cualquier geodésica a infinito. Intuitivamente más o menos corresponde con la propiedad de que se puede extender todas las geodésicas arbitrariamente. Si un espacio no es geodésicamente completo, significa que algunas geodésicas terminan en un punto específico, posiblemente porque acaban en una singularidad, o porque el espaciotiempo tiene un borde con que se choca la geodésica (véase el caso a en la Figura 10.3).

Una siguiente pregunta que uno podría hacer es si un espaciotiempo dado tiene una superficie de Cauchy, es decir una superficie espacial, que podría servir para especificar las condiciones iniciales, de modo que toda la evolución posterior esté determinada por estas condiciones iniciales. Si uno encuentra una superficie Σ que sea espacial en todos los puntos y especifica en esta superficie las posiciones y velocidades iniciales de las partículas y flujos, no está garantizado que para cualquier espaciotiempo esta superficie Σ sea una superficie de Cauchy. La geometría del

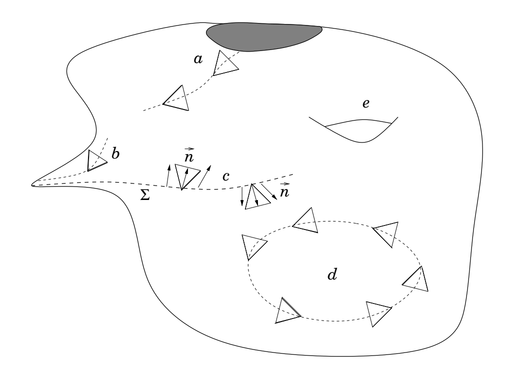


Figura 10.3: *Estructuras no-locales en las soluciones de las ecuaciones de Einstein: un espacio es geodésicamente completo si cualquier geodésica se extiende hasta el infinito, sino acaba en un borde o una singularidad (a). No en todos los espacios es posible definir superficies de Cauchy, ya que una influencia puede venir del infinito sin haber cruzado la superficie espacial Σ (b). Un espacio es orientable con respecto al tiempo, si hay una definición global de la dirección en que fluye el tiempo (c). Curvas temporales cerradas son curvas que son temporales en todos sus puntos, pero que vuelven al mimso punto y permiten viajar al pasado (d). Las ecuaciones de Einstein determinan sólo la geometría local, pero no dicen nada sobre la topología de la solución (e).*

espaciotiempo puede ser así que, aún especificando condiciones iniciales en Σ, haya curvas temporales que vengan del infinito y que no crucen la superficie Σ. En otras palabras, en algunos espacios es posible que un evento en el futuro de la superficie Σ esté influenciado por hechos no especificados en las condiciones iniciales (véase el caso b en la Figura 10.3). En tales espacios no se puede calcular el futuro del espacio entero, sino solo de trozos pequeños en el futuro inmediato de Σ. El espacio de anti-De Sitter es un ejemplo de un espacio donde no existen superficies de Cauchy.

Otro ejemplo de una propiedad geométrica curiosa es la orientabilidad del tiempo. En algunos espacios, no es posible definir un campo vectorial temporal que tenga la misma orientación en todo el espacio. Intuitivamente más o menos se corresponde al hecho de que no haya una dirección global en la que fluya el tiempo, sino que en distintas regiones del espacio el tiempo fluye en direcciones diferentes. Concretamente, si en tal espacio se coge una foliación con superficies Σ_k que son espaciales en todos sus puntos, el campo vectorial n^μ ortogonal a Σ_k es temporal en todos los sitios. En espacios que no son orientables con respecto al tiempo resulta que, aunque se elija n^μ siempre dirigido hacia el futuro, la orientación de n^μ con respecto a Σ_k no es la misma en todo el espacio. En algunas regiones n^μ estará "por un lado" de Σ_k y en otras regiones por "el otro lado" (véase caso c en la figura 10.3). En cada punto del espacio tenemos una estructura lorentziana bien definida, con una dirección temporal bien definida, pero globalmente no es posible definir un sentido global para una dirección temporal.

Un último ejemplo de una estructura geométrica contraintuitiva, en este caso incluso patológica, es la presencia de curvas temporales cerradas. Ya hemos mencionado que una curva temporal puede servir como posible trayectoria de una partícula masiva. En principio es posible que la variación de la orientación de los conos de luz de punto en punto sea tal que existan curvas temporales cerradas, es decir, que sean temporales a lo largo de toda la curva y además vuelvan al mismo punto (véase el caso d en la Figura 10.3). Estas curvas temporales cerradas en general no son geodésicas, pero en principio sería posible que en estos espacios un observador viajara a su propio pasado. En general se suelen descartar estas soluciones no físicas, por los probemas lógicos relacionados con poder influenciar el pasado, pero no hay nada en las ecuaciones de Einstein que impida estas soluciones. Un ejemplo es el universo de Gödel.

Finalmente, merece la pena enfatizar que las ecuaciones de Einstein determinan la métrica $g_{\mu\nu}$

y por lo tanto la geometría local del espacio, pero no determinan la topología de las soluciones, que es una propiedad global. La topología del espacio viene determinada (entre otras cosas) por las condiciones de contorno, cosa que no entra en las ecuaciones de Einstein, sino que hay que poner a mano. Dos espacios pueden tener la misma estructura geométrica, pero una topología distinta. Un ejemplo muy sencillo es el caso del espacio plano. Tanto el espacio de Minkowski cuadrimensional $\mathbb{R}^{1,3}$

$$ds^2 = dt^2 - dx^2 - dy^2 - dz^2 \tag{10.49}$$

como el producto directo del espacio de Minkowski bidimensional con el toro $\mathbb{R}^{1,1} \times T^2$

$$ds^2 = dt^2 - dx^2 - R_1^2 d\theta^2 - R_2^2 d\varphi^2, \tag{10.50}$$

con R_1 y R_2 los radios del toro y $\theta, \varphi \in [0, 2\pi]$, son soluciones de las ecuaciones del vacío y ambas son planas, ya que $R_{\mu\nu\rho\lambda} = 0$. A escalas $\ell \ll R_1, R_2$, un observador no es capaz de ver la diferencia entre estas dos soluciones, ya que en cada punto se puede hacer el cambio de coordenadas local $y = R_1\theta$, $z = R_2\varphi$, de modo que las métricas de las dos soluciones son la misma. Sin embargo la topología es muy diferente. Por ejemplo, cualquier curva en $\mathbb{R}^{1,3}$ se puede contraer a un punto, mientras que una curva enroscada a lo largo de θ en $\mathbb{R}^{1,1} \times T^2$ no es contraíble. En general es difícil determinar la estructura topológica (como la presencia de asas o ciclos no-triviales, véase el caso e en la Figura 10.3) a base de la métrica $g_{\mu\nu}$.

Capítulo 11

Los tests clásicos de la relatividad general

Si la Teoría de la Relatividad resulta correcta, en Alemania dirán que soy alemán, en Suiza que soy suizo y en Francia que soy un gran científico. Si la teoría es falsa, para los franceses seré un suizo, para los suizos un alemán y para los alemanes un judio.
(Atribuido a A. Einstein)

En los últimos dos capítulos hemos desarollado una teoría moderna de la gravedad, a base de sólo tres principios: el Principio de Equivalencia, el Principio de Covariancia y el Principio de Mínimo Acoplo. Estos tres principios sencillos y elegantes llevan directamente a las ecuaciones de Einstein, que forman el núcleo de la teoría de la relatividad. Pero el hecho de que una teoría sea obtenida de manera elegante, a base de primeros principios, no implica que sea correcta, es decir, que describa bien la Naturaleza. Al fin y al cabo, lo que determina si una teo'ria física es correcta o no, son los resultados expermentales. Por lo tanto, para tener un poco más de confianza en la teoría, habrá que pasar por unas cuantas pruebas experimentales.

11.1. El límite newtoniano

Cuando Einstein en 1915 presentó su versión final de la relatividad general, no había ninguna razón experimental para dudar de la validez de la teoría newtoniana de la gravedad, salvo quizá el desplazamiento del perihelio de Mercurio (véase la sección 11.2). Al contrario, entre 1841 y 1845 los astrónomos John Couch Adams (1819-1892) en Cambridge y Urbain Le Verrier (1811-1877) en París utilizaron independientemente la teoría de la gravedad newtoniana para predecir correctamente la posición del en ese momento desconocido planeta Neptuno a base de las irregularidades en el movimiento de Urano, un hecho que entonces fue recibido como un gran éxito tanto para la investigación teórica como para la gravedad newtoniana. Por lo tanto, si la relatividad general pretende ser una teoría moderna de la gravedad, que corrige a la teoría newtoniana, más vale que por lo menos recupere ésta en algún límite no-relativista.

El límite adecuado resulta ser el límite cuando el campo gravitatorio es débil y (casi) estático, es decir, cuando la curvatura del espaciotiempo es prácticamente cero y todas las partículas (incluso las que causan el campo gravitatorio) tienen velocidades mucho más bajas que la velocidad de la luz, $v \ll 1$. Con esta última condición estamos eligiendo implícitamente un sistema de coordenadas específico, el sistema donde todas las velocidades son pequeñas. Aunque de este modo estamos perdiendo covariancia general, no nos importa demasiado de momento, ya que la teoría

newtoniana presupone un sistema de coordenadas preferido (un espacio y un tiempo absoluto).[1]

La condición del campo gravitatorio débil permite escribir la métrica como una perturbación del espacio de Minkowski

$$g_{\mu\nu} \;=\; \eta_{\mu\nu} \;+\; \varepsilon\, h_{\mu\nu}, \tag{11.1}$$

con $\varepsilon \ll 1$ el parámetro de expansión. Tomando en cuenta que hasta primer orden en ε, la métrica inversa toma la forma

$$g^{\mu\nu} \;\approx\; \eta^{\mu\nu} \;-\; \varepsilon\, h^{\mu\nu}, \tag{11.2}$$

en primera aproximación los símbolos de Christoffel y el tensor y el escalar de Ricci vienen dados por (ejerc)

$$\begin{aligned}
\Gamma^{\rho}_{\mu\nu} \;&\approx\; \frac{1}{2}\varepsilon\, \eta^{\rho\lambda}\Big[\partial_\mu h_{\lambda\nu} + \partial_\nu h_{\mu\lambda} - \partial_\lambda h_{\mu\nu}\Big], \\
R_{\mu\nu} \;&\approx\; \frac{1}{2}\varepsilon\, \partial^2 h_{\mu\nu}, \qquad\qquad R \;\approx\; \frac{1}{2}\varepsilon\, \partial^2 h,
\end{aligned} \tag{11.3}$$

donde $h = \eta^{\mu\nu} h_{\mu\nu}$ es la traza de la perturbación de la métrica, $\partial^2 = \eta^{\mu\nu}\partial_\mu\partial_\nu$ es el d'alambertiano en el espacio de Minkowski y donde hemos fijado los grados de libertad no-físicos de $h_{\mu\nu}$ a través de la elección de gauge

$$\partial^\mu h_{\mu\nu} - \frac{1}{2}\partial_\nu h = 0. \tag{11.4}$$

En primera aproximación, las ecuaciones de Einstein para un campo gravitatorio débil se convierten entonces en

$$\frac{1}{2}\varepsilon\Big[\partial^2 h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}\partial^2 h\Big] \;\approx\; -\kappa T_{\mu\nu}. \tag{11.5}$$

También resultará útil la versión de esta ecuación sin traza, es decir, la aproximación para campo débil de (10.22),

$$\frac{1}{2}\varepsilon\partial^2 h_{\mu\nu} \approx -\kappa\Big[T_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}T\Big], \tag{11.6}$$

donde $T = \eta^{\mu\nu}T_{\mu\nu}$. Obsérvese que la ecuación de Einstein (11.6) toma la forma de una ecuación de ondas inhomegénea para $h_{\mu\nu}$. Discutiremos esta propiedad en más detalle en en Capítulo **??**.

La segunda condición, que todas las velocidades consideradas sean no-relativistas, implica que podemos utilizar el mismo parámetro de expansión ε para el movimiento de las partículas, ya que $v \sim \varepsilon \ll 1$. En particular, el tiempo propio (5.25) de una partícula moviéndose en un campo gravitatorio es por lo tanto en primera aproximación igual al tiempo en el sistema de referencia:

$$d\tau \;\approx\; dt. \tag{11.7}$$

Una partícula que se mueve con velocidad $v \sim \varepsilon$ viaja una distancia dx^i en el intervalo de tiempo dt de modo que

$$dx^i \;\sim\; \varepsilon\, dt, \tag{11.8}$$

y por lo tanto para cualquier función $f(x^i, t)$ vemos que

$$\partial_t f \;\sim\; \varepsilon\, \partial_i f. \tag{11.9}$$

En otras palabras, la derivada de una función con respecto al tiempo es un orden en ε mayor que el gradiente y pueden ser despreciada en comparación con éstas.

Con estas aproximaciones, no sólo podemos ver que las ecuaciones de Einstein se reducen a la ecuación de movimiento de una partícula en un potencial gravitatorio, sino que también

[1] Aquí vemos que en realidad el tiempo absoluto de Newton no es más que la coordenada temporal del sistema de referencia donde todas las partículas se mueven a velocidades no relativistas y el espacio absoluto a las superficies t constantes en estas coordenadas.

podemos afirmar que la constante de proporcionalidad κ en (10.20) realmente es $\kappa = 8\pi G_N$, como dijimos en la seccion 10.2.

El potencial newtoniano se saca de la ecuación de la geodésica (7.42), la ecuación de movimiento de una partícula libre en relatividad general. Las componentes espaciales se reducen en nuestra aproximación a

$$\frac{d^2x^i}{dt^2} + \Gamma^i_{tt} + \Gamma^i_{jk}\frac{dx^j}{dt}\frac{dx^k}{dt} \approx 0, \tag{11.10}$$

donde hemos utilizado la aproximación (11.7). El último término se puede despreciar frente a los primeros dos, ya que también es de orden ε^2, debido a (11.8). Por último tenemos que en primera aproximación

$$\Gamma^i_{tt} \approx \frac{1}{2}\varepsilon\,\eta^{ij}\Big[2\partial_t h_{tj} - \partial_j h_{tt}\Big] \approx -\frac{1}{2}\varepsilon\,\eta^{ij}\partial_j h_{tt} \approx -\frac{1}{2}\partial^i g_{tt}, \tag{11.11}$$

donde en la segunda igualdad hemos despreciado las derivadas temporales frente al gradiente por (11.9) y en la tercera igualdad hemos utilizado la expansión (11.1) para escribir el resultado en término de la métrica $g_{\mu\nu}$. La ecuación de la geodésica se reduce por lo tanto a

$$\frac{d^2x^i}{dt^2} \approx \frac{1}{2}\partial^i g_{tt}. \tag{11.12}$$

Comparando esta expresión con la segunda ley de Newton para una partícula en un campo gravitatorio (5.57) y tomando en cuenta que $\partial^i = -(\vec{\nabla})^i$, vemos que podemos identificar en primera aproximación el potencial de Newton Φ con la componente g_{tt} de la métrica:[2]

$$g_{tt} \approx 1 + 2\Phi. \tag{11.13}$$

Efectivamente, en el Capítulo 12 veremos que la solución de las ecuaciones de Einstein que describe un objeto masivo estático con masa m y con simetría esférica viene dada por la métrica de Schwarzschild,

$$ds^2 = \Big(1 - \frac{2G_N m}{r}\Big)dt^2 - \Big(1 - \frac{2G_N m}{r}\Big)^{-1}dr^2 - r^2\Big(d\theta^2 + \sin^2\theta d\varphi^2\Big), \tag{11.14}$$

donde G_N es la constante de Newton, de modo que con (11.13) recuperamos la forma conocida del potencial newtoniano para objectos esféricos

$$\Phi \approx -\frac{G_N m}{r}. \tag{11.15}$$

Efectivamente, con la identificación del potencial newtoniano (11.15), la ecuación de la geodésica (11.12) se convierte en la conocida segunda ley de Newton para una partícula en un potencial gravitatorio: $\ddot{\vec{x}} = -\vec{\nabla}\Phi$. Obsérvese que aquí vemos concretamente lo que ya habíamos dicho en el Capítulo 9: lo que para Newton es una fuerza gravitatoria F_g que actúa a distancia, es para Einstein un término geométrico, la componente Γ^i_{tt} de la conexión de Levi-Civita del espaciotiempo en que se mueve la partícula.

El valor de la constante de proporcionalidad κ lo obtenemos a través de las ecuaciones de Einstein (11.6) para un campo débil, originado por una distribución de materia fría (polvo) con velocidades no relativistas. En esta aproximación la única componente no cero del tensor de energía-momento (10.1) es $T^{tt} \approx \rho_0$, de modo que bajando los índices y tomando la traza tenemos que

$$T_{tt} \approx \rho_0, \qquad\qquad T \approx \rho_0. \tag{11.16}$$

[2]Nótese que asumimos que el espaciotiempo sea asintóticamente plano (para la definición véase Capitulo 12). Si el espaciotiempo no es asintóticamente plano, la definición del potencial gravitatorio es un poco más complicada.

La componente (tt) de la ecuación de Einstein (11.6) es por lo tanto

$$\varepsilon\Delta h_{tt} \;\approx\; \kappa\rho_0, \tag{11.17}$$

donde hemos despreciado las derivadas temporales ∂_t^2 frente al laplaciano $\Delta = \partial_i\partial^i$. De (11.1) y (11.13) sabemos que en primera aproximación podemos identificar εh_{tt} con dos veces el potencial gravitatorio, de modo que (11.17) se convierte en

$$\Delta\Phi \;\approx\; \frac{1}{2}\,\kappa\rho_0. \tag{11.18}$$

Una comparación con la ecuación de Poisson (5.58) fija la constante de proporcionalidad de (10.20) en $\kappa = 8\pi G_N$.

En resumen, tanto la Segunda Ley de Newton (5.57) como la ecuación de Poisson (5.58) están contenidas dentro de la relatividad general, en el límite donde el campo gravitatorio es débil y todas las velocidades consideradas son mucho más bajas que la velocidad de la luz, es decir, cuando tanto los efectos de la curvatura del espacio como los efectos de la relatividad especial son despreciables. Este es efectivamente el rango de validez que uno esperaría para la mecánica y la gravedad newtoniana.

11.2. El perihelio de Mercurio

Einstein propuso tres tests para comprobar su relatividad general: el avance del perihelio de Mercurio, la deflexión de la luz por un objeto masivo y la dilatación gravitacional del tiempo, conocidos hoy como los *tres tests clásicos*. Estos tests son más fáciles en la presencia de un objeto masivo estático con simetría esférica. Como ya hemos mencionado en la sección anterior, la métrica que describe un espaciotiempo de tal objeto es la métrica de Schwarzschild,

$$ds^2 = \Big(1 - \frac{2M}{r}\Big)dt^2 - \Big(1 - \frac{2M}{r}\Big)^{-1} dr^2 - r^2\Big(d\theta^2 + \sin^2\theta d\varphi^2\Big), \tag{11.19}$$

donde la constante $M = G_N m$, con dimensión **L**, es una medida para la masa m del objeto. Para el Sol, M corresponde aproximadamente a 1,5 km (para más detalles sobre la derivación de esta métrica y su significado físico referimos al Capítulo 12).

El test del avance del perihelio de Mercurio consiste en que la relatividad general predice una corrección relativista a la primera ley de Kepler en las órbitas de los planetas: en lugar de que el planeta siga órbitas perfectamente elípticas (en ausencia de otros planetas), la trayectoria tiene en primera aproximación la forma de un elipse que gira lentamente en el plano de la órbita, de modo que el perihelio, la posición más cercana al Sol, se mueve cierto ángulo $2\pi\varepsilon$ a lo largo de un círculo con respecto a su perihelio anterior (véase Figura 11.1)

Para entender bien el argumento relativista, es conveniente repasar brevemente el cálculo newtoniano. Es interesante observar que ambos cálculos llegan a las mismas ecuaciones (salvo correcciones relativistas) por argumentos totalmente distintos.

En el formalismo newtoniano, la manera más fácil de calcular las órbitas de las planetas es a partir de la ley de conservación de momento angular L y la ley de conservación de energía E para una partícula con masa m_0 (con $m_0 \ll M/G_N$) en un potencial newtoniano $V(r) = -m_0 M/r$,

$$m_0 r^2\dot{\varphi} = L, \qquad\qquad \frac{1}{2}m_0(\dot{r}^2 + r^2\dot{\varphi}^2) - \frac{m_0 M}{r} = E, \tag{11.20}$$

donde hemos asumido, sin pérdida de generalidad, que la partícula se mueve en el plano ecuatorial $\theta = \pi/2$. Sustituyendo $\dot{\varphi}^2$ en la ley de conservación de energía por su valor en función de

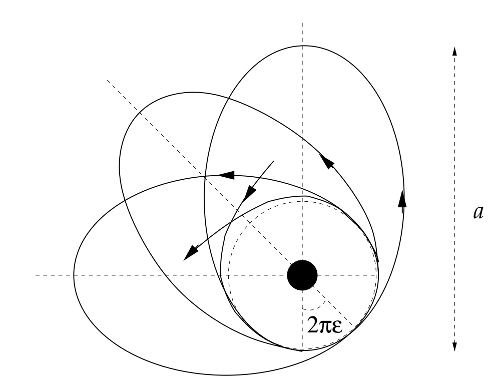


Figura 11.1: *El avance del perihelio (exagerado): Cada vuelta alrededor del Sol, el perihelio se mueve con un ángulo* $2\pi\varepsilon$

L y cambiando a la variable $u = r^{-1}$, la ley de conservación de energía se convierte en (ejerc.)

$$\Big(\frac{du}{d\varphi}\Big)^2 + u^2 = \frac{2m_0E}{L^2} + \frac{2m_0^2M}{L^2}u, \tag{11.21}$$

donde hemos escrito la variable u como función de φ a través de la regla de la cadena

$$\frac{du}{dt} = \frac{du}{d\varphi}\frac{d\varphi}{dt}. \tag{11.22}$$

La ecuación (11.20) se puede integrar directamente, pero es más fácil si la derivamos otra vez con respecto a φ, para obtener la ecuación de Binet (ejerc.)

$$\frac{d^2u}{d\varphi^2} + u = \frac{m_0M}{L^2}, \tag{11.23}$$

cuya solución son las conocidas secciones cónicas

$$u \;=\; \frac{m_0M}{L^2}\Big(1 + e\;\cos\varphi\Big), \tag{11.24}$$

donde $e = \sqrt{1 + 2L^2E/m_0^2M}$ es la excentricidad de la órbita. Para energía $E < 0$, la excentricidad e es menor que 1 y la trayectoria es un elipse.

Miramos ahora el caso relativista. Para calcular la órbita de una partícula con masa m_0 en relatividad general tenemos que calcular las geodésicas temporales en la solución de Schwarzschild (11.19). Las componentes t y φ de la ecuación de la geodésica (7.46) y la condición (7.53) de que sea temporal en este espacio vienen dadas por

$$\begin{aligned} &\Big(1 - \frac{2M}{r}\Big)\dot{t} = k, \qquad\qquad r^2\dot{\varphi} = \frac{L}{m_0}, \\ &\Big(1 - \frac{2M}{r}\Big)\dot{t}^2 - \Big(1 - \frac{2M}{r}\Big)^{-1}\dot{r}^2 \;-\; r^2\dot{\varphi}^2 = 1, \end{aligned} \tag{11.25}$$

donde k y L son constantes de integración. Por comparación con (11.20) podemos interpretar L como el momento angular de la partícula, mientras k es la constante de proporcionalidad entre la coordenada temporal t y el tiempo propio de la partícula. Sustituyendo las primeras dos ecuaciones en la última obtenemos (ejerc.)

$$\Big(\frac{dr}{d\varphi}\Big)^2 \;+\; r^2\Big(1 - \frac{2M}{r}\Big)\Big[1 + \frac{m_0^2r^2}{L^2}\Big] \;-\; \frac{k^2m_0^2r^4}{L^2} \;=\; 0, \tag{11.26}$$

y cambiando a la variable $u = 1/r$, la ecuación se reduce a (ejerc.)

$$\Big(\frac{du}{d\varphi}\Big)^2 \ + \ u^2 \ = \ \frac{m_0^2(k^2-1)}{L^2} \ + \ \frac{2m_0^2M}{L^2}u \ + \ 2Mu^3, \tag{11.27}$$

La ecuación (11.27) es muy parecida a la ley de conservación de energía (11.21) en el caso newtoniano, salvo el último término. Esto implica que podemos interpretar k como una medida para la energía de la partícula,

$$E = \frac{1}{2}m_0(k^2-1), \tag{11.28}$$

mientras que el último término, $2Mu^3$, actúa como una perturbación a la ecuación newtoniana. Efectivamente, el tercer término es mucho más pequeño que el segundo puesto que

$$\frac{3Mu^2}{m_0^2M/L^2} \ \approx \ \frac{3L^2}{a^2m_0^2} \ = \ \frac{3M(1-e^2)}{a}, \tag{11.29}$$

donde a es el eje mayor del elipse y donde en la última igualdad hemos utilizado la segunda y la tercera ley de Kepler,

$$T = \frac{2\pi a^2\sqrt{1-e^2}}{L}, \qquad\qquad T^2 = \frac{2\pi^2a^3}{M}. \tag{11.30}$$

Dado que para el Sol M es aproximadamente 1,5 km y para Mercurio $a = 6\cdot 10^6$ km y $e = 0,2056$, vemos que, para el caso de Mercurio, el término relativista en (11.27) es del órden de 10^{-7} veces menor que el segundo término. Obsérvese que para los demás planetas, el eje mayor es aún más grande y la excentricidad es más pequeña (salvo para Plutón), de modo que la perturbación es aún menor.

En lugar de intentar resolver la ecuación (11.27), resulta más fácil derivarla con respecto a φ y obtener la llamada ecuación relativista de Binet (ejerc.)

$$\frac{d^2u}{d\varphi^2} \ + \ u \ = \ \frac{m_0^2M}{L^2} \ + \ 3Mu^2, \tag{11.31}$$

donde otra vez el último término es una pertubación a la ecuación de Binet (11.23). Definiendo el parámetro de perturbación adimensional

$$\varepsilon = \frac{3m_0^2M^2}{L^2} = \frac{3M}{a(1-e^2)}, \tag{11.32}$$

cuyo valor para Mercurio es del orden de 10^{-7}, y suponiendo que la solución es en primera aproximación de la forma $u = u_0 + \varepsilon u_1$, con u_0 la solución newtoniana (11.24), encontramos que la ecuación para la perturbación en primer orden viene dada por (ejerc.)

$$\frac{d^2u_1}{d\varphi^2} \ + \ u_1 \ \approx \ \frac{m_0^2M}{L^2}\Big[\Big(1+\frac{e^2}{2}\Big) \ + \ 2e\cos\varphi \ + \ \frac{1}{2}e^2\cos 2\varphi\Big]. \tag{11.33}$$

Sustituyendo el Ansatz

$$u_1 = A + B\varphi\cos\varphi + C\cos 2\varphi \tag{11.34}$$

en la ecuación (11.33), encontramos que existe una solución para (ejerc.)

$$A = \frac{m_0^2M}{L^2}\Big(1+\frac{e^2}{2}\Big), \qquad B = \frac{m_0^2Me}{L^2}, \qquad C = -\frac{m_0^2Me^2}{6L^2}, \tag{11.35}$$

de modo que la solución hasta primer orden viene dada por (ejerc.)

$$u \ \approx \ \frac{m_0^2M}{L^2}\Big(1+e\ \cos\varphi\Big) \ + \ \varepsilon\frac{m_0^2M}{L^2}\Big[1+e\varphi\sin\varphi + e^2\Big(\frac{1}{2}-\frac{1}{6}\cos 2\varphi\Big)\Big]. \tag{11.36}$$

De los términos de corrección, el segundo es el más importante porque es proporcional a φ y representa por lo tanto un efecto acumulativo. Despreciando los otros términos y quedándonos sólo con términos hasta orden ε, podemos escribir (11.36) como (ejerc.)

$$u \approx \frac{m_0^2 M}{L^2}\Big[1 + e\ \cos\Big((1-\varepsilon)\varphi\Big)\Big]. \tag{11.37}$$

Esta ecuación es muy parecida a la del elipse (11.24), sólo que hay una perturbación de orden ε en la dependencia angular. Por lo tanto, la curva (11.37) representa en primera aproximación un elipse, pero el periodo de la trayectoria no es de 2π, sino un poco mayor,

$$\frac{2\pi}{1-\varepsilon} \approx 2\pi(1+\varepsilon). \tag{11.38}$$

En otras palabras, el planeta no sigue una trayectoria perfectamente elíptica, sino que el elipse se va girando un poco cada revolución, de modo que el planeta alcanza su perihelio un poquito más tarde cada vez. El retraso viene dado precisamente por el parámetro ε. La definición (11.32) de ε nos enseña que el efecto es independiente de la masa m_0 del planeta, pero que aumenta cuanto más cerca esté el planeta del Sol y más excéntrica sea su órbita.

El desplazamiento del perihelio de Mercurio de 574 arcosegundos por siglo ya era conocida desde la primera mitad del siglo XIX. En los 1850, Le Verrier, el co-descubridor de Neptuno, calculó que la mayor parte del efecto medido (531 arcosegundos por siglo) es debido a la influencia gravitational de los demás planetas, sobre todo de Venus. Pero el cálculo de Le Verrier también mostró que aún quedaba un desplazamiento adicional de unos 42 arcosegundos por siglo sin explicar, mucho más que el error observacional, lo que hizo que el problema era considerado como una de las anomalías más grandes de la mecánica newtoniana. Le Verrier mismo creó que este efecto era debido a la existencia de un planeta interior a Mercurio, bautizado Vulcano.[3] El problema de esta teoría era que para poder causar el efecto calculado, Vulcano debería ser tan grande que sería perfectamente visible desde la Tierra. A pesar de que se anuciaron algunas observaciones (después de la publicación de Le Verrier), su existencia nunca se ha llegado a confirmar y el problema permaneció hasta que Einstein calculó en 1915 las correcciones relativistas al potencial newtoniano y vió que causaban un desplazamiento al perihelio de Mercurio adicional al efecto newtoniano, que cuadraba con las observaciones. Efectivamente el efecto relativista es de $42,98 \pm 0,04$ arcosegundos por siglo, con un error teórico perfectamente compatible con las observaciones.

11.3. La deflexión de la luz

El cálculo de la deflexión de la luz es muy parecido al cálculo del perihelio de Mecurio, sólo que, en vez de calcular las geodésicas temporales en la solución de Schwarzschild, hay que calcular las geodésicas nulas. Por lo tanto, restringiéndonos al movimiento plano, las componentes t y θ de la ecuación de la geodésica y la condición de que éste sea nula imponen las siguientes ecuaciones, análogas a (11.25),

$$\Big(1-\frac{2M}{r}\Big)\dot{t} = k, \qquad\qquad r^2\dot{\varphi} = \ell,$$
$$\Big(1-\frac{2M}{r}\Big)\dot{t}^2 - \Big(1-\frac{2M}{r}\Big)^{-1}\dot{r}^2 - r^2\dot{\varphi}^2 = 0, \tag{11.39}$$

donde ahora el punto indica una derivación con respecto al parámetro σ de la geodésica nula, que ya no tiene la interpretación del tiempo propio y ℓ es el momento angular asociado a la geodésica.

[3] No confundir con el planeta paterno de Mr. Spock en Star Trek. Aunque sí hay relación: cuando Gene Roddenberry (1921 - 1991) inventó Star Trek en 1964, recicló el nombre del planeta de Le Verrier para los primeros extraterrestres que aparecen en su serie.

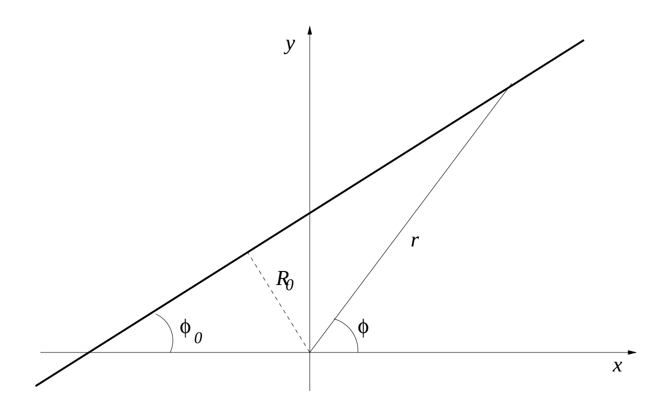


Figura 11.2: *La ecuación de la recta en* $\mathbb{R}^2$ *en coordenadas polares* (r, φ)*: las constantes de integración* R_0 *y* φ_0 *representan respectivamente la distancia de la recta hasta el origen y el ángulo con respecto al eje* x*.*

Con una derivación idéntica al caso de las geodésicas temporales, obtenemos el análogo de la ecuación (11.27) (donde otra vez $u = 1/r$),

$$\Big(\frac{du}{d\varphi}\Big)^2 \;+\; u^2 \;=\; \frac{k^2}{\ell^2} \;+\; 2Mu^3, \tag{11.40}$$

y derivando con respecto a φ la ecuación relativista de Binet para geodésicas nulas,

$$\frac{d^2u}{d\varphi^2} \;+\; u \;=\; 3Mu^2. \tag{11.41}$$

Obsérvese que tanto el momento angular ℓ como la medida para la energía k han desaparecido de la ecuación. Al tratarse de partículas sin masa, que se mueven siempre con la velocidad de la luz, estas cantidades están en realidad determinadas por la distancia r al centro de gravedad.

Hay un límite en que esta ecuación es fácil de resolver de manera exacta: cuando $M = 0$, la solución de Schwarzschild (11.19) se reduce al espacio de Minkowski in coordenadas esféricas y en esta caso la solución de (11.41) viene dada por

$$u(\varphi) \;=\; R_0^{-1}\sin(\varphi - \varphi_0). \tag{11.42}$$

En la expresión en términos de r,

$$r\sin(\varphi - \varphi_0) = R_0, \tag{11.43}$$

reconocemos la ecuación de la recta en $\mathbb{R}^2$ en coordenadas polares, donde R_0 representa la distancia de la recta hasta el origen y φ_0 el ángulo con respecto al eje x (véase Figura 11.2), lo que efectivamente es el resultado esperado para una geodésica nula en el espacio de Minkowski.

Igual que para el perihelio de Mercurio, intentaremos resolver la ecuación de Binet (11.41) en el límite donde el término a la derecha es una perturbación del resultado newtoniano, sólo que ahora nuestro parámetro adimensional viene dado por

$$\varepsilon = \frac{3M}{R_0}. \tag{11.44}$$

Exigir por lo tanto que $\varepsilon \ll 1$ es considerar sólo geodésicas cuya distancia R_0 hasta el centro del potencial gravitatorio sea mucho más grande que M, es decir, geodésicas que pasan lejos del objeto en el centro.[4] En este caso la ecuación de Binet (11.41) se puede escribir como

$$\frac{d^2u}{d\varphi^2} \;+\; u \;=\; \varepsilon R_0 u^2, \tag{11.45}$$

[4]Como comparación: en el Capítulo 12 veremos que el horizonte del agujero negro neutro y estático está localizado en el radio de Schwarzschild $r = 2M$.

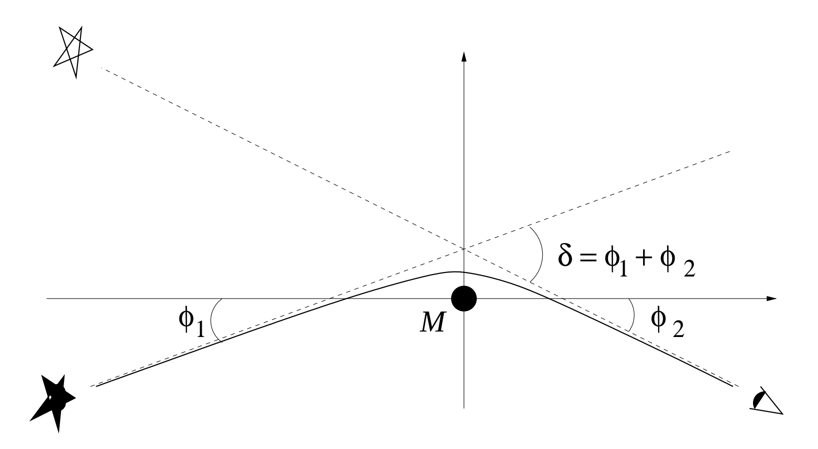


Figura 11.3: *La deflexión de la luz: la luz de un objeto (estrella negra) que pasa cerca de un objeto masivo, está desviada por el campo gravitatorio y proyecta una imagen (estrella blanca) en una posición distinta.*

y asumiendo que la solución $u(\varphi)$ es una perturbación del resultado newtoniano

$$u(\varphi) \;=\; u_0(\varphi) \;+\; \varepsilon u_1(\varphi), \tag{11.46}$$

con $u_0(\varphi)$ dado por (11.42), se puede ver fácilmente que la ecuación para u_1 es de la forma

$$\frac{d^2u_1}{d\varphi^2} \;+\; u_1 \;=\; R_0^{-1}\sin^2(\varphi-\varphi_0), \tag{11.47}$$

cuya solución viene dada, hasta primer orden en ε, por

$$u_1(\varphi) \;\approx\; \frac{1}{3}R_0^{-1} \;+\; \frac{1}{3}\beta R_0^{-1}\cos(\varphi-\varphi_0) \;+\; \frac{1}{3}R_0^{-1}\cos^2(\varphi-\varphi_0), \tag{11.48}$$

con β una constante adimensional arbitraria, que podemos poner a cero para el caso que nos interesa. La solución completa, hasta primer orden en ε, es por lo tanto

$$u(\varphi) \approx R_0^{-1}\sin(\varphi-\varphi_0) \;+\; \frac{M}{R_0^2}\Big[1 \;+\; \cos^2(\varphi-\varphi_0)\Big], \tag{11.49}$$

o, en terminos de r,

$$r\sin(\varphi-\varphi_0) \;\approx\; R_0 \;-\; \frac{M}{R_0}\Big[1 \;+\; \cos^2(\varphi-\varphi_0)\Big], \tag{11.50}$$

donde efectivamente se puede ver el término entre corchetes como una perturbación de (11.43).

Lejos del centro del objeto masivo (es decir para $r \ggg M$), la solución tiende asintóticamente al espacio de Minkowski, de modo que para $r \to \infty$, la trayectoria de la luz tiende a una recta. Sólo cerca del objeto masivo, la desviación es apreciable y la trayectoria real interpola entre una recta entrante y otra saliente. Consecuentemente, el ángulo δ de la deflexión de la luz es el ángulo entre estas dos rectas, que viene dado por la suma de los ángulos φ_1 y φ_2 que hacen las rectas con el eje x (véase Figura 11.3). Para determinar δ en términos de M y R_0, basta con expresar φ_1 y φ_2 en función de éstas, a través de (11.50).

Cuando $u \to 0$, $\varphi \to \varphi_2$ a lo largo de la recta saliente, de modo que (11.49) se reduce a

$$0 \;\approx\; R_0^{-1}\sin(\varphi-\varphi_2) \;+\; \frac{M}{R_0^2}\Big[1+\cos^2(\varphi-\varphi_2)\Big], \tag{11.51}$$

o, en hasta primer orden en $\varphi-\varphi_2$,

$$\varphi_2 \;\approx\; \varphi \;+\; \frac{2M}{R_0}. \tag{11.52}$$

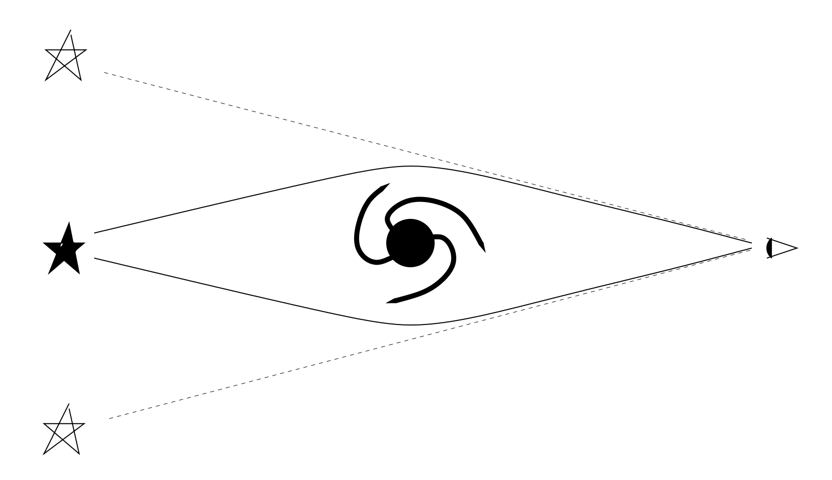


Figura 11.4: *Una lente gravitatoria: un objeto lejano (estrella negra) proyecta multiples imágenes (estrellas blancas) en el cielo terrestre, debido que un objeto masivo más cercano enfoca la luz emitida en direcciones diferentes.*

Por otro lado a lo largo de la recta entrante $\varphi \to \varphi_1 + \pi$ asintóticamente, de modo que (11.50) impone la condición

$$\varphi_2 \approx -\varphi + \frac{2M}{R_0}, \tag{11.53}$$

donde el signo viene del hecho de que $\sin(\varphi_1 + \pi) = -\sin\varphi_1$.

El ángulo de la deflexión de la luz es por lo tanto

$$\delta = \varphi_1 + \varphi_2 \approx \frac{4M}{R_0}. \tag{11.54}$$

El efecto de la deflexión de la luz hace que la trayectoria de la luz que ha pasado cerca del centro de la solución de Schwarzschild se desvia y proyecta una imagen en una parte del cielo distinta a su verdadera posición (Véase Figura 11.3). Concretamente para el Sol (para el cual, como ya hemos dicho $M \approx 1,5\,km$), el ángulo de deflexión de una estrella es de unos 1,7 segundos de arco, si la luz pasa justo fuera del radio del Sol. En circunstancias normales esto no se notaría, ya que la propia luz del Sol taparía la de la estrella, pero el efecto sí podría manifestarse durante los eclipses solares.

En 1919, dos expediciones inglesas intentaron medir el efecto en el eclipse solar del 29 de mayo de ese año en Brasil y en la isla de Príncipe, delante de la costa de África occidental, ésta última liderada por el renombrado astrofísico británico Arthur Eddington (1881 - 1944). Debido a las dificultades técnicas inherentes a un experimento de tal calibre (en realidad demasiadas para en la época), la calidad de los resultados era muy pobre, con poca precisión y pocos datos para hacer una buena estadística de error. Aún así el equipo en Principe medío una deflexión entre 1,31 y 1,91 segundo de arco y el equipo en Brasil una media de 1,91 segundo de arco con un aparato y de 0,86 con otro, un resultado que parece confirmar cualitativamente las prediciones de relatividad general, aunque claramente no cuantitativamente (¡siendo la predicción newtoniana 0,8!). Incluso hay constancia de que Eddington, un gran admirador de Einstein y de la relatividad general, fue un tanto selectivo con el análisis de los datos y los métodos estadísticos empleados. Pero más por el peso de la autoridad de Eddington que por los argumentos científicos, los resultados de las expediciones de 1919 fueron recibidos como la gran confirmación de la relatividad general y dió Einstein su estatus de prototipo de científico famoso.

Sin embargo, hoy en día ya nadie duda de que la deflexión de la luz sea un efecto real, gracias a la cosmología observacional. El efecto de la deflexión de la luz no es típico de la solución de

Schwarzschild (sólo es más fácil calcularlo aquí, debido a la simetría de la solución), sino que ocurre en general cerca de cualquier conjunto grande de masa y energía. Hay numerosos ejemplos de galaxias y cuásares que están en posiciones (ligeramente) distintas en el cielo, pero muestran exactamente el mismo espectro, indicando que en realidad se trata del mismo objeto. La luz de estos objetos lejanos, inicialmente emitida en direcciones (ligeramente) distintas, está enfocada al pasar cerca de otro objeto muy masivo más cercano. En este caso el objecto cercano realmente actúa como una *lente gravitacional*, proyectando multiples imágenes en el cielo terrestre (véase Figura 11.4). Este efecto ocurre en tres dimensiones, por lo que puede dar lugar a configuraciones como la *Cruz de Einstein* (el cuásar $QSO2237 + 0305$, que proyecta 5 imágenes en forma de cruz) o (secciones de) los *anillos de Einstein* que aparecen en las fotografías de campo profundo, hechas por el Hubble Space Telescope. La distorsión de los objetos en estas imágenes nos puede dar información sobre la cantidad y la distribución de la materia oscura en el universo.

El mismo efecto (aunque aún no observado) ocurre cerca de agujeros negros: el campo gravitatorio de estos objetos es tan grande que la luz puede dar completamente la vuelta o incluso llegar a orbitar el agujero negro. Los efectos de distorsión gravitacional de las imágenes son mayores cuanto más uno se acerca (y no suele ser considerada en imágenes artísticas de agujeros negros, por lo que éstas son poco fiables).

11.4. El efecto Doppler gravitacional

En la sección 9.3 ya habíamos deducido la existencia de un efecto Doppler gravitacional a base del Principio de Equivalencia e incluso habíamos derivado unos primeros resultado cualitativos para un campo gravitatorio constante a través de un ascensor acelerado. En esta sección queremos hacer la derivación con más cuidado con las herramientas de relatividad general y geometría diferencial. Sin embargo, el estudiante atento notará un paralelismo entre esta derivación y la de la sección 9.3.

El efecto Doppler gravitacional no sólo ocurre para campos gravitacionales constantes, sino para cualquier espaciotiempo estático. En grandes líneas, un espaciotiempo estático quiere decir que existe un sistema de coordenadas donde podemos dividir las cuatro coordenadas en una temporal, digamos t, y tres espaciales x^i y que las componentes de la métrica $g_{\mu\nu}$ no dependen de la coordenada temporal t. Además la simetría $t \rightarrow -t$ de las soluciones estáticas impide que la métrica tenga términos cruzados g_{ti}, de modo que la forma más general de una métrica estática viene dada por

$$ds^2 \;=\; g_{tt}(x^k)\, dt^2 \;-\; g_{ij}(x^k)\, dx^i\, dx^j. \tag{11.55}$$

Consideremos ahora dos observadores, llamados el emisor y el detector, en reposo en dos posiciones diferentes del espaciotiempo, ambos equipados con un reloj atómico. Sean $x_e^\mu = x_e^\mu(\tau_e)$ y $x_d^\mu = x_d^\mu(\tau_d)$ las lineas de universo de cada observador, parametrizados por el tiempo propio de cada uno. El emisor emite con cierte regularidad señales hacia el detector, donde el intervalo $d\tau_e$ entre dos emisiones viene dado por

$$d\tau_e^2 \;=\; g_{\mu\nu}(x_e)\, dx^\mu\, dx^\nu \;=\; g_{tt}(x_e)\, dt^2. \tag{11.56}$$

Por otro lado, el intervalo entre dos detecciones $d\tau_d$ no necesariamente tiene que coincidir con $d\tau_e$, de modo que

$$d\tau_d^2 \;=\; (\gamma\, d\tau_e)^2 \;=\; g_{tt}(x_d)\, dt^2, \tag{11.57}$$

donde γ es el factor de dilatación temporal debido al campo gravitatorio y podemos obtener una expresión para γ dividiendo ambas expresiones,

$$\gamma \;=\; \sqrt{\frac{g_{tt}(x_d)}{g_{tt}(x_e)}}. \tag{11.58}$$

En otras palabras, si el periodo de entre la emisión de dos señales consecutivas es T_e, entonces el periodo entre dos detecciones consecutivas es

$$T_d \;=\; T_e \sqrt{\frac{g_{tt}(x_d)}{g_{tt}(x_e)}}, \tag{11.59}$$

o, en términos de frecuencias,

$$\nu_d \;=\; \nu_e \sqrt{\frac{g_{tt}(x_e)}{g_{tt}(x_d)}}. \tag{11.60}$$

Por lo tanto, si $g_{tt}(x_e) \,<\, g_{tt}(x_e)$, tenemos que $\nu_d \,<\, \nu_e$. En otras palabras, un rayo de luz perderá energía al intentar salir del pozo potencial gravitatorio y sufrirá un corrimiento hacia el rojo. Nótese, sin embargo, que no hemos particularizado a ninguna solución específica, de modo que el efecto es no sólo válido para la solución de Schwarzschild (11.19), sino en general para cualquier solución estática.

La fórmula (11.59) es una generalización de la derivación de la sección 9.3, hecha a base del Principio de Equivalencia. De la sección 11.1 sabemos que en espaciotiempos estáticos el potencial newtoniano Φ está relacionado con la componentes g_{tt} de la métrica a través de (11.13), de modo que podemos aproximar (11.59) por

$$T_d \;\approx\; T_e \sqrt{\frac{1+2\Phi(x_d)}{1+2\Phi(x_e)}} \;\approx\; T_e \Big[1+\Phi(x_d)-\Phi(x_e)\Big], \tag{11.61}$$

donde en la última igualdad hemos expandido la raíz cuadrada en un desarrollo de Taylor hasta primer orden en Φ, con lo que recuperamos la formula (9.17).

En 1960 los físicos americanos Robert Pound (1919 - 2010) y Glen Rebka (1931 -) comprobaron experimentalmente el efecto en el Jefferson Laboratory de la Universidad de Harvard. La idea central del experimento está basada en la emisión y absorción de fotones cuando un átomo pasa de un estado excitado a su estado base y vice versa. Dado que la energía de la luz emitida por los átomos exitados al decaer al estado cambia al viajar por el campo gravitacional de la Tierra, los fotones ya no pueden excitar los átomos similares en el detector. Para poder cuantificar el efecto, Pound y Rebka hicieron mover la fuente con una velocidad v con respecto al detector para causar un adicional efecto Doppler (cinemático)

$$\nu_d \;=\; \nu_e \sqrt{\frac{1-v}{1+v}}, \tag{11.62}$$

que compensaría el efecto Doppler gravitacional. Sólo con una velocidad v específica, los dos efectos se contrarrestarían completamente, maximizando la probabilidad de absorción de los fotones. En particular, para el experimento de Pound y Rebka, había buscar v tal que (suponiendo que el campo gravitatorio de la Tierra viene descrita por la métrica de Schwarzschild)

$$\sqrt{\frac{1-v}{1+v}\cdot\frac{1-\frac{2M}{R_0+h}}{1-\frac{2M}{R_0}}} \;=\; 1, \tag{11.63}$$

donde R_0 es el radio de la Tierra y h la altura de la torre del Jefferson Laboratory, que medía 22,5 m de altura. El cambio de energía en la onda electromagnética debido a esta diferencia de potencial gravitatoria es del orden de $\delta E/E \approx 2,5\cdot 10^{-15}$, lo que les obligó a usar rayos gamma de 14 keV y aprovechar el recién descubierto efecto Mösbauer (1958) para absorber el retroceso del núcleo del átomo al emitir un fotón tan energético. Los resultados originales de Pound y Rebka confirmaron las predicciones de la relatividad general con una exactitud de un $10\,\%$, mejorándolos unos años más tarde al $1\,\%$. Los experimentos actuales miden una precisión de un $0,01\,\%$.

Otro experimento que confirmó la dilatación temporal gravitacional fue el Gravity Probe A, un satélite lanzado en 1976 vérticulamente hasta una altura de 10.000 km, donde según las predicciónes teóricas un reloj debería correr $4,5 \cdot 10^{-8}\,\%$ más rápido que en la superficie de la Tierra. Equipado con un máser de hidrógeno de alta precisión y comparándolo con otro máser en la superficie, Gravity Probe A era capaz de medir que la variación en el ritmo de los relojes concordaba con las predicciones teóricas con una precisión de $7 \cdot 10^{-5}\,\%$.

Pero quizá la aplicación más famosa del efecto Doppler gravitacional se encuentra en el sistema de GPS (Global Positioning System). Para determinar su posición sobre la superficie terrestre, el aparato GPS tiene que contactar mediante señales electromagnéticas con por lo menos 4 de los 24 satélites que formas la red GPS, que orbitan a Tierra a una altura de 20.000 km. Comparando los tiempos recibidos de los diferentes satélites y triangulando los resultados, un aparato GPS (civil) es capaz de conseguir una precisión de menos de 10 m en cuestion de segundos. Para obtener la precisión deseada, los satélites están equipados con unos relojes atómicos que emiten con una frecuencia de $1,023 \cdot 10^6 Hz$ para uso civil y $10,23 \cdot 10^6 Hz$, donde el error en cada emisión no puede superar los 30 ns $(30 \cdot 10^{-9} s)$.

Sin embargo, al estar en órbita a 20.000 km de altura, hay que tomar en cuenta dos efectos relativistas: el primero es el efecto de la dilatacion temporal cinemática de relatividad especial, debido a su velocidad (orbital) con respecto al aparato GPS en la superficie terrestre, que hace que los relojes atómicos se atrasan con respecto a la Tierra. Por otro lado está el efecto de relatividad general, debido a la diferencia de potencial gravitatoria que hace que los relojos atómicos corren más de prisa arriba que abajo. Los dos efectos se contrarrestan y el resultado final depende de la altura de los satélites. Cuanto menor la altura h sobre la superficie terreste, más alta es la velocidad orbital y más pequeña la diferencia en potencial gravitatoria, de modo que más dominará el efecto de relatividad especial. Dos dos efectos se cancelan cuando $h = \frac{1}{2}R_0$, en caso de la Tierra a una altura de 3200 km, pero a los 20.000 km de altura donde están colocados los satélites GPS, el efecto dominante es el de relatividad general. Concretamente, su velocidad orbital de 14.000 km/h hace retrasar los relojos 7 μs al día, pero la diferencia de altura los adelanta con 45 μs al día, resultando en un efecto de 38 μs al día, tres órdenes de magnitud más que el error permitido (30 ns) para alcanzar la precisión exigida.

Al diseñar el sistema GPS, los ingenieros han tenido que tomar en cuenta este efecto, entre otras cosas instalando relojes atómicos con una frecuencia ligeramente más pequeña (en la Tierra), de modo que en órbita tendría exactamente la frecuencia deseada. Al no tomar en cuenta el efecto Doppler gravitacional, el sistema GPS tendría un error acumulativo (!) de unos 10 km al día, volviendo el sistema GPS completamente inútil en cuestión de horas. El efecto Doppler gravitacional en el sistema GPS es probablemente el único efecto de relatividad general que se manifiesta en la vida cotidiana.

11.5. El efecto Shapiro

A los tres testsc lásicos de relatividad general, propuestos por Einstein, se suele añadir un cuarto, el efecto del *retraso temporal gravitacional*, o el *efecto Shapiro*, propuesto por el astrofísico americano Irwin Shapiro (1929 -) en 1964. La idea consiste en mandar una señal de radar a un planeta cercano, digamos Venus, cuando éste está (casi) alineado con la Tierra y el Sol, midiendo cuánto tarde la señal en reflectar en la superficie y ser detectada otra vez en la Tierra. La relatividad general predice que al viajar por el espaciotiempo curvado por el Sol, la trayectoria de la señal es más larga y el tiempo necesario para recorrerla mayor que si el espacio fuera plano (véase Figura 11.5).

Al igual que en el caso de la deflexión de la luz, tenemos que tratar con geodésicas nulas en la solución de Schwarzschild, cuya aproximación hasta primer orden en M/R_0 ya hemos calculado en (11.50). Sin embargo, como ya consideramos el espaciotiempo hasta orden M/r, podemos

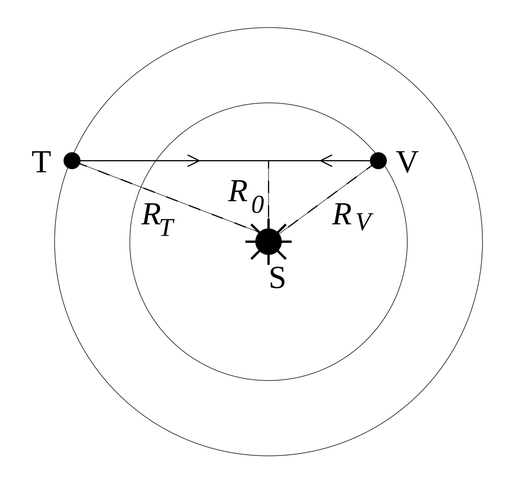


Figura 11.5: *El efecto Shapiro: Un rayo de luz (linea continua) emitida desde la Tierra (T) refleja en la superficie de Venus (V) y está detectado otra vez en la Tierra. Al viajar por el campo gravitatorio del Sol (S), la trayectoria por el espaciotiempo curvo es más larga y la señal sufrirá un retraso en comparación con el espacio plano.*

aproximar las geodésicas por la expresion (11.43) a orden cero. La condición de que la geodésica sea nula viene dada por

$$\Big(1-\frac{2M}{r}\Big)dt^2-\Big(1-\frac{2M}{r}\Big)^{-1}dr^2\ -\ r^2d\varphi^2=0, \tag{11.64}$$

donde podemos eliminar $d\varphi^2$ a través de la expresión de la geodésica no perturbada. Diferenciando (11.43) obtenemos que

$$\sin\varphi\,dr\ +\ r\cos\varphi\,d\varphi=0, \tag{11.65}$$

o, equivalentemente, usando la misma expresión (11.43) (ejerc.),

$$r^2\,d\varphi^2\ =\ \frac{R_0^2}{r^2-R_0^2}\,dr^2, \tag{11.66}$$

con R_0 la distancia mínima a la que se acerca la señal del origen. Sustituyendo esto en (11.64), tenemos que (ejerc.)

$$\begin{aligned} dt^2 &= \Big[\Big(1-\frac{2M}{r}\Big)^{-2}\ +\ \Big(1-\frac{2M}{r}\Big)^{-1}\frac{R_0}{r^2-R_0^2}\Big]dr^2 \\ &\approx \frac{r^2}{r^2-R_0^2}\Big[1\ +\ \frac{4M}{r}\ -\ \frac{2MR_0^2}{r^3}\Big]dr^2, \end{aligned} \tag{11.67}$$

donde en la segunda ecuación nos hemos quedado con los términos de primer orden del desarrollo de Taylor en M/r. Sacando la raíz cuadrada de esta expresión, otra vez hasta primer orden en M/r,

$$dt\ \approx\ \frac{\pm\,r}{\sqrt{r^2-R_0^2}}\Big[1\ +\ \frac{2M}{r}\ +\ \frac{MR_0^2}{r^3}\Big]\,dr, \tag{11.68}$$

podemos obtener una expresión para dt en función de r, cuya integral corresponde al tiempo que tarda la señal, integrando esta última expresión entre la posicion de ambos planetas. Obsérvese que el signo menos corresponde a la parte de la trayectoria entre la posición R_T de la Tierra y R_0 y el signo positivo con la parte entre la posición R_V de Venus y R_0. El tiempo total para la señal

para llegar a Venus es por lo tanto

$$
\begin{aligned}
T &\approx -\int_{R_T}^{R_0} \frac{r}{\sqrt{r^2 - R_0^2}} \left[1 + \frac{2M}{r} + \frac{MR_0^2}{r^3}\right] dr + \int_{R_0}^{R_V} \frac{r}{\sqrt{r^2 - R_0^2}} \left[1 + \frac{2M}{r} + \frac{MR_0^2}{r^3}\right] dr \\
&= \sqrt{R_T^2 - R_0^2} + \sqrt{R_V^2 - R_0^2} + M\left[\sqrt{1 - \frac{R_0^2}{R_T^2}} + \sqrt{1 - \frac{R_0^2}{R_V^2}}\right] \\
&\qquad + 2M \log\left[\left(\frac{R_T}{R_0} + \sqrt{\frac{R_T^2}{R_0^2} - 1}\right)\left(\frac{R_V}{R_0} + \sqrt{\frac{R_V^2}{R_0^2} - 1}\right)\right]. \qquad (11.69)
\end{aligned}
$$

Los primeros dos términos representan la distancia entre la Tierra y Venus si el espaciotiempo fuera plano, y los términos proporcionales a M son las correcciones relativistas debidas a la métrica de Schwarzschild .

El efecto Shapiro para los planetas internos (Mecurio, Venus y Marte) es del orden de 200 μs y fue confirmado por Shapiro mismo en 1967 con una precisión de un $3\,\%$ y de un $1\,\%$ en experimentos posteriores. Sin embargo la era de los viajes interplanetarios ha permitido obtener resultados muchos mejores, midiendo el retraso en señales en la comunicación con naves espaciales como los Mariner 6 y 7 cerca del Sol y el Viking en Marte con una precisión de $0,1\,\%$ y en el 2003 una precisión de $0,0012\,\%$ con el Cassini, orbitando Saturno.

11.6. Tests modernos de la relatividad general

Aparte de los tests clásicos y sus variantes modernos, hay varios otras pruebas a las que se ha sometido la relatividad general en las últimas décadas.

Entre 2004 y 2005 el satélite Gravity Probe B, orbitando la Tierra, ha estado recogiendo datos para comprobar dos predicciones de la relatividad general, hasta entonces no comprobadas: el efecto geodético y el efecto de Lense-Thirring. El *efecto geodético*, también llamado el *efecto De Sitter*, fue derivado en 1916 por el físico y astrónomo holandés Willem De Sitter (1872 - 1934) al estudiar correcciones relativistas al movimiento del sistema Tierra-Luna. El efecto geodético predice la precesión del momento angular de una masa de prueba en órbita alrededor de una masa central, como en la solución de Schwarzschild (11.14), debido al transporte paralelo de este vector por el espacio curvo. Por otro lado, el *efecto Lense-Thirring*, descubierto por los físicos austriacos Josef Lense (1890 - 1985) y Hans Thirring (1888 - 1976) en 1918, predice una precesión adicional (ortogonal al efecto geodético, pero más pequeño) del momento angular de la masa de prueba, si el objeto central está rotando (por ejemplo la solución de Kerr, discutida en el Capítulo **??**). El efecto Lense-Thirring está íntimamente ligado con el efecto de arrastre de sistemas inerciales (discutido en detalle en la sección **??**), que afirma que un objeto masivo en rotación arrastra consigo el espaciotiempo en su alrededor, como una bola rotando en un fluido viscoso.

El experimento de Gravity Probe B consistía de cuatro giróscopos superconductores, mantenidos constantemente a una temperatura menor de 2 K por helio superfluido para minimizar las interferencias y orientados a IM Pegasi, una estrella de referencia a 329 años-luz en la constelación de Pegaso. La precisión de estos giróscopos era tal que eran capaces de medir variaciones en la orientación del eje de rotación menores que 0,5 miliarcosegundos por año (al ángulo bajo el que se vería el grosor de un pelo humano desde 32 km de distancia!).

Las predicciones teóricas para una masa de prueba orbitando la Tierra son de $-$6606,1 miliarcosegundos por año para el efecto geodético y de $-$39,2 milisegundos de arco por año para el efecto Lense-Thirring. Los resultados experimentales fueron respectivamente de $-$6601,8 $\pm$ 18,3 y $-$37,2 $\pm$ 7,2 milisegundos de arco por año, una confirmación del efecto geodético con una

precisión del $0,5\,\%$. La precisión en la medición del efecto Lense-Thirring es bastante más baja, debido al hecho de que la magnitud del efecto era de la mismo orden que ruido que no se había tomado en cuenta al modelar el experimento. Varios años de análisis y extracción de datos aún han permitido alcanzar una precisión de un $19\,\%$.

Todos los tests de la relatividad general que hemos discutido hasta ahora han sido tests hechos dentro del sistema solar y por lo tanto, por muy exitosos que hayan sido, sólo han podido comprobar el régimen de campo débil de la teoría, es decir, donde las correcciones a la teoría newtoniana son pequeñas. Mucho más difícil, pero por lo menos igual de importante es comprobar la teoría en situaciones donde el campo gravitatorio es muy fuerte. Para eso el estudio de los púlsares binarios es idoneo.

Un púlsar binario es un púlsar[5] que forma parte de un sistema estelar binario, junto con una enana blanca, una estrella de neutrones y ocasionalmente incluso otro púlsar. En general una estrella de neutrones por si sola ya produce un campo gravitatorio muy fuerte, puesto que contiene aproximadamente una masa solar comprimida en un radio de unos diez kilómetros. Pero si además un objeto de estas dimensiones está orbitando junto con otro acompañante masivo vertiginosamente su centro de masa común a una distancia del orden de distancia Tierra-Luna, la relatividad general predice que este sistema es una fuente constante de ondas gravitacionales. Estas ondas gravitacionales no son detectables directamente (aún), pero sí se puede comprobar su existencia en las variaciones de la órbita del pulsar: las ondas gravitacionales se llevan una parte de la energía potencial y cinética del sistema, disminuyendo el periodo de órbita y haciendo que las masas se acercan lentamente hasta chocarse.

En 1974 los astrónomos americanos Russell A. Hulse (1950 -) y Joseph Taylor (1941 -) descubieron el primer púlsar binario, PSR B1913+16, a 21.000 años-luz en la constelación del Águila. El sistema consiste de un púlsar y una estrella de neutrones, ambos con una masa de aproximadamente 1,4 masas solares, a una distancia que varía entre 750.000 y 3.000.000 km, unos pocos radios solares. Las dos estrellas tienen un periodo orbital de poco menos de 8 horas. Desde su descubrimiento, hace más de 35 años, Hulse y Taylor han seguido observando PSR B1913+16, encontrando que el periodo orbital ha bajado unos 76,5 μs por año, y el semi-eje mayor ha diminuido 3,5 m por año. Estos resultados confirman las predicciones teóricas hasta un $0,2\,\%$. En 1993 se les otorgó a Hulse y Taylor el Premio Nobel de Física por sus resultados.

En 2003, la radioestronoma italiana Marta Burgay (1976 -) descubrió un púlsar doble, PSR J0737-3039, un sistema binario formado por dos púlsares con masas alrededor de 1,3 masas solares y un periodo orbital de sólo 2,4 horas. La importancia de PSR J0737-3039 es el púlsar binario con el periodo más corto conocido, lo que permite hacer pruebas de efectos relativistas con muy alta precisión. También aquí se mide un pérdida de energía en forma de ondas gravitacionales, resultando en que la órbita disminuye con 7 mm al día, confirmando también las predicciones teóricas con alta precisión.

Más general, la relatividad general predice la existencia de agujeros negros, ondas gravitacionales y un universo dinámico, cuyas propiedades discutiremos en detalle en la Parte **??**. Aunque los primeros dos no han sido detectados directamente, sí hay indicaciones indirectas muy fuertes de su existencia. Para el caso de las ondas gravitacionales, ya hemos mencionado la pérdida de energía en forma de radiación gravitatoria en los sistemas de púlsares binarios. Y para el caso de los agujeros negros, se han observado los efectos sobre la materia cercana de campos gravitatorios tan fuertes, que ningún modelo astrofísico permite que sea causado por otra cosa que agujeros negros. Quiza el ejemplo más convincente es Sagitario A*, una fuente de radioondas localizada

[5]Los púlsares son estrellas de neutrones que emiten pulsos de radiación electromagnética muy regulares en la dirección de la Tierra. Las estrellas de neutrones, los restos de una estrella que han sobrevivido una explosión de supenova, son objetos tremendamente compactos y tienen campos magnéticos muy fuertes, que hace que emiten radiación electromagnética fuerte a lo largo del eje magnético. Si el eje magnético no está alineado con el eje de rotación, el haz de radiación gira como un faro por el espacio. Si este haz pasa por la Tierra, se observa la estrella de neutrones como un púlsar. El periodo de pulsaciónes es muy estable y varía de púlsar a púlsar entre segundos y milisegundos.

en el centro de nuestra galaxia. Monitorizaciones de los movimientos de las estrellas cercanas han revelado que en la vecindad de Sagitario A* hay una concentración de unos 4 millones de masas solares en un radio de menos de 44 millones de kilómetros (menos que el radio de la órbita del planeta Mercurio), lo que sólo se puede interpretar como un agujero negro supermasivo. Aún así, las indicaciones de la existencia de estos fenómenos es solamente indirecta. La detección directa de sobre todo ondas gravitacionales sería no sólo una confirmación importante de la relatividad general, sino también abriría una nueva ventana para entender la dinámica de la gravedad y la estructura del universo.

Por otro lado sí tenemos confirmaciones más directas de las implicaciones cosmológicas de las ecuaciones de Einstein, aunque aquí las predicciones no sólo van por cuenta de la relatividad general, sino también de la física de partículas. Las ecuaciones de Einstein permiten un amplio espectro de soluciones cosmológicas, sin precisar cuál de ellas realmente corresponde con nuestro universo, pero el paradigma cosmológico sugerido por la teoría, en combinación con las actuales conocimientos de física de partículas, ha permitido a lo largo de la segunda mitad del Siglo XX reproducir con bastante precisión ciertas observaciones cosmológicas, como la abundancia de hidrógeno y helio en el universo y la temperatura del fondo cósmico de microondas. Pero gracias a las observaciones precisas desde los años 1990, se han podido determinar los parámetros cosmológicos con tanta precisión que ya podemos hablar de un Modelo Estándar Cosmológico, en analogía con el Modelo Estándar de la física de partículas.

Pero también es verdad que es en la cosmología donde quedan más preguntas abiertas. Una de las cosas que hemos aprendido recientemente es que la materia observada directamente (es decir en forma de estrellas, planetas, etc) sólo contribuye un $4,6\,\%$ del contenido total de energía en el universo. Las observaciones de la distribución de materia en galaxias y cúmulos de galaxias relevan que hay 5 veces más masa de la que vemos ópticamente. Hay fuertes indicaciones de que esta *materia oscura*, que sólo podemos observar a través de sus efectos gravitatorios, no es bariónica (es decir, no está hecha de protones y neutrones, como la materia que conocemos). Pero incluso el conjunto de materia sólo forma una fracción del contenido total de energía del universo: más preciso un $27,6\,\%$ ($4,6\,\%$ visible y $23\,\%$ materia oscura). Los restante $71\,\%$ aparecen en una forma desconocida, denominada *energía oscura* y es la responsable de la observada expansión acelerada del universo. Las observaciones cosmológicas sugieren que esa energía oscura se comporta muy parecido a una constante cosmológica, aunque es demasiado pronto para excluir del todo otras posibilidades. En otras palabras, ¡menos del $5\,\%$ del contenido del universo viene en forma conocida, mientras el $95\,\%$ es de un tipo que nunca ha sido detectado en un laboratorio!

Se suele interpretar la materia oscura como lagunas en los actuales modelos de física de partículas y se espera que por ejemplo la supersimetría pueda proporcionar candidatos para la materia oscura (aunque hay también intentos de modificar las ecuaciones de Einstein, generando un comportamiento distinto de la gravedad a escalas cósmicas, que explicaría los efectos observados sin tener que recurrir a la materia de un tipo desconocido). Por otro lado, el problema de la energía oscura es de otro tipo, ya que en este caso ni siquiera sabemos de qué forma de energía estamos hablando.

Si la energía oscura realmente resultara ser una constante cosmológica Λ (de todas las posibilidades, la más familiar), seguiría la pregunta de qué representa esta Λ. La teoría cuántica de campos predice una energía del vacío en forma de la energía del punto cero de los campos cuánticos, que resultaría en una constante cosmológica efectiva. Sin embargo esta constante cosmológica sería 120 (!) órdenes de magnitud mayor que la energía oscura de nuestro universo. La pregunta por lo tanto no es sólo qué representa la energía oscura observada, sino también por qué la energía del vacío es tan baja en comparación con las predicciones de la teoría cuántica de campos.

En realidad nos encontramos aquí en el umbral de un terreno completamente desconocido: la gravedad cuántica. Toda la relatividad general (y la gran mayoría de la física explicada en este curso) es puramente clásica y no toma en cuenta el caracter cuántico de la Naturaleza. Hay

resultados importantes en el régimen semi-clásico, como la teoría cuántica de campos en espacios curvos, que trata el contenido de energía y materia como cuántico, pero la gravedad misma como clásica, por ejemplo en la cercancía de agujeros negros. Pero no tenemos una una descripción intrínsicamente cuántica de la interacción gravitatoria, en parte porque el régimen donde estos efectos cuánticos son importantes están completamente fuera de alcance de nuestra tecnología experimental actual: se supone que la interacción gravitatoria entre dos protones es del mismo orden que la electromagnética a la escala de Planck,[6]

$$m_P = \sqrt{\frac{\hbar}{G_N}} \sim 10^{19}\, GeV, \tag{11.70}$$

(en unidades con $c = 1$, sino tendríamos que $m_P = \sqrt{\hbar c/G_N}$) unos 15 órdenes de magnitud más que las energías producidas en el LHC. Aunque hay algunos candidatos, como la teoría de cuerdas y gravedad cuántica de bucles (loop quantum gravity), los intentos de construir estas teorías son puramente teóricos, basados principalmente en elegancia y consistencia matemática. Por eso, la física cerca de agujeros negros y la constante cosmológica pueden ser unas de las pocas pistas que tenemos para echar algo de luz sobre este campo.

[6] $1\, GeV = 10^9\, eV$. Un electronvoltio ($1\, eV$) es la energía que adquiere un electrón cuando pasa por la diferencia de potencial de un voltio. Un electronvoltio con aproximadamente $1,6 \cdot 10^{-19} J$.

Parte IV

Soluciones de las Ecuaciones de Einstein

Capítulo 12

La solución de Schwarzschild

No esperaba que se pudiera obtener una solución exacta de manera tan sencilla.
(A. Einstein, en una carta a K. Schwarzschild)

Ya hemos dicho en varias ocasiones que las ecuaciones de Einstein son muy difíciles de resolver, debido a su caracter no lineal, de modo que la superposición de dos soluciones no es una nueva solución. No es difícil entender la razón física para esta no linealidad: sabemos que el espacio se curva debido a su contenido de masa y energía. Pero la propia curvatura del espaciotiempo contiene energía, de modo que la misma curvatura es una fuente de curvatura. En otras palabras, la gravedad no sólo se acopla a la energía y la materia, sino también a sí misma, lo que resulta en ecuaciones no lineales.

Einstein mismo creyó inicialmente que sus ecuaciones eran tan complicadas que nunca se encontraría una solución exacta. Sin embargo, pocos meses después de la publicación de la relatividad general, en 1916, Karl Schwarzschild (1873 - 1916) halló la solución exacta de un objeto estático con simetría esférica y en los últimos 90 años decenas, si no cientos de soluciones exactas han sido encontradas. En este capítulo discutiremos esta solución de Schwarzschild y su significado físico.

12.1. La derivación de la solución de Schwarzschild

La solución de Schwarzschild es una solución estática de las ecuaciones del vacío, con simetría esférica. Por lo tanto es una buena descripción para el campo gravitatorio causado por objetos masivos esféricos, como estrellas y planetas. En particular, son precisamente las geodésicas de la métrica de Schwarzschild que nos permite calcular correcciones relativistas a las órbitas planetarias y la deflexión de la luz, como hemos visto en el Capítulo 11.

En ausencia de energía y materia, el tensor de energía-momento $T_{\mu\nu}$ es cero y (la parte sin traza de) las ecuaciones de Einstein se reducen a

$$R_{\mu\nu} = 0. \tag{12.1}$$

A primera vista puede resultar extraño que las ecuaciones del vacío admiten soluciones no-triviales, aparte del espacio de Minkowski. Pero justo gracias a la no linealidad existe una gran clase de soluciones no-triviales del vacío. Las métricas que satisfacen la ecuaciones del vacío (12.1) se llaman *Ricci-planas*. Las soluciones que son Ricci-planas en general no son planas, pero sí son soluciones del vacío (en ausencia de una constante cosmológica).

La solución de Schwarzschild es la solución no-trivial más sencilla, debido a su gran cantidad

de simetría. Para empezar es estática, lo que significa en grandes líneas que no hay evolución en el tiempo y por lo tanto existe un sistema de coordenadas en que la métrica es independiente de la coordenada temporal. Además no puede haber términos cruzados del tipo $g_{ti}\, dtdx^i$, ya que la presencia de estos términos rompería la invariancia $t \to -t$ de las soluciones estáticas. Luego, la simetría esférica implica que el espacio tiene una simetría $SO(3)$, es decir, que es invariante bajo rotaciones ortogonales en tres dimensiones.[1] Existen por lo tanto unas coordenadas angulares θ y φ tales que las secciones espaciales $t = t_0$ se puedan escribir como

$$ds^2 \;=\; -f(r)dr^2 \;-\; r^2\Big(d\theta^2 +\; \sin^2\theta d\varphi^2\Big), \tag{12.2}$$

donde r es la coordenada radial. En principio el factor delante de la parte angular puede ser una función arbitraria $h^2(r)$, pero con un cambio de coordenadas $\tilde{r} = h(r)$ siempre se puede escribir la métrica en la forma (12.2).

Sabemos que las ecuaciones de Einstein son demasiado difíciles de resolver directamente, pero gracias a la simetría de la solución de Schwarzschild podemos escribir un *Ansatz* (una propuesta) de la forma

$$ds^2 \;=\; e^{2A(r)}dt^2 \;-\; e^{2B(r)}dr^2 - r^2\Big(d\theta^2 +\; \sin^2\theta d\varphi^2\Big), \tag{12.3}$$

con $A(r)$ y $B(r)$ dos funciones que quedan por determinar. Nótese cómo el Ansatz refleja la simetría esférica y el hecho de que la solución es estática en la ausencia de términos cruzados y el hecho de que A y B son funciones de r únicamente. La idea ahora es sustituir este Ansatz en las ecuaciones del vacío (12.1) para determinar A y B.

No es difícil calcular los símbolos de Christoffel. Los símbolos no-triviales vienen dados por (ejerc.):

$$\begin{array}{lll}
\Gamma_{tt}^{r} = e^{2(A-B)}A', & \Gamma_{r\theta}^{\theta} = r^{-1}, & \Gamma_{\theta\varphi}^{\varphi} = \operatorname{cotg}\theta, \\
\Gamma_{tr}^{t} = A', & \Gamma_{r\varphi}^{\varphi} = r^{-1}, & \Gamma_{\varphi\varphi}^{r} = -r\sin^2\theta\, e^{-2B}, \\
\Gamma_{rr}^{r} = B', & \Gamma_{\theta\theta}^{r} = -re^{-2B}, & \Gamma_{\varphi\varphi}^{\theta} = -\sin\theta\cos\theta,
\end{array} \tag{12.4}$$

de modo que las componentes del tensor y el escalar de Ricci que no son cero son (ejerc.)

$$\begin{aligned}
R_{tt} &= -e^{2(A-B)}\Big[A'' + (A')^2 - A'B' + 2r^{-1}A'\Big], \\
R_{rr} &= A'' + (A')^2 - A'B' - 2r^{-1}B', \\
R_{\theta\theta} &= e^{-2B}\Big[rA' - rB' + 1\Big] - 1, \\
R_{\varphi\varphi} &= \sin^2\theta\, R_{\theta\theta}, \\
R &= -2e^{-2B}\Big[A'' + (A')^2 - A'B' + 2r^{-1}(A' - B') + r^{-1}\Big] + 2r^{-1}
\end{aligned} \tag{12.5}$$

donde la prima denota la derivada con respecto a r. Igualar todas las componentes del tensor de Ricci a cero, como exige las ecuaciones de Einstein, nos hace resolver un sistema de 4 ecuaciones diferenciales no-lineales acopladas para dos incógnitas. En principio este sistema está sobredeterminado y no tiene soluciones, pero veremos que no todas las ecuaciones son independientes.

Multiplicando R_{tt} por $e^{-2(A-B)}$ y sumándolo con R_{rr} obtenemos que

$$0 \;=\; e^{-2(A-B)}R_{tt} + R_{rr} \;=\; -2r^{-1}(A' + B'), \tag{12.6}$$

[1]Ojo, esto no implica que las secciones espaciales sean $\mathbb{R}^3$. La solución de Schwarzschild misma ya es un contraejemplo, como veremos en breve.

de modo que $B(r) = -A(r) + c_0$. La constante de integración c_0 no tiene significado físico, ya que se puede absorber en una redefinición de la coordenada temporal $t' = e^{c_0}t$. Sin pérdida de generalidad podemos poner $c_0 = 0$ y la solución general de (12.6) es

$$A(r) = -B(r). \tag{12.7}$$

Sustituyendo esta condición en la ecuación para $R_{\theta\theta}$ encontramos

$$0 \;=\; e^{2A}\Big[2rA' + 1\Big] - 1 \;=\; \Big[re^{2A}\Big]' - 1. \tag{12.8}$$

Este ecuación se puede integrar directamente

$$e^{2A} \;=\; 1 - \frac{2M}{r}, \tag{12.9}$$

donde $-2M$ es una constante de integración con dimension de longitud **L**, elegida de esta forma para futura conveniencia.

Vemos por lo tanto que resolviendo dos ecuaciones (la de $R_{\theta\theta}$ y una combinación de R_{tt} y R_{rr}) hemos determinado las funciones $A(r)$ y $B(r)$ por completo, salvo una constante de integración. Estará claro de (12.5) que la ecuación para $R_{\varphi\varphi}$ está satisfecha si la de $R_{\theta\theta}$ lo está. Queda por lo tanto sólo otra combinación de las ecuaciones para R_{tt} y R_{rr} por resolver. Lo más sencillo es la misma ecuación para R_{rr}. No es difícil averiguar que efectivamente (12.9) satisface idénticamente la ecuación

$$A'' \;+\; 2(A')^2 \;+\; 2r^{-1}A' \;=\; 0, \tag{12.10}$$

para cualquier valor de M.

La solución de Schwarzschild viene por lo tanto dada por

$$ds^2 \;=\; \Big(1 - \frac{2M}{r}\Big)dt^2 \;-\; \Big(1 - \frac{2M}{r}\Big)^{-1}dr^2 \;-\; r^2 d\Omega_2^2, \tag{12.11}$$

donde $d\Omega_2^2 = d\theta^2 + \sin^2\theta d\varphi^2$ es el elemento de línea de la dos-esfera $\mathbb{S}^2$.

Discutamos ahora la interpretación y algunas propiedades físicas de la solución de Schwarzschild. Por razones que se harán claras en breve, nos limitaremos de momento a la región donde $r > 2M$. En la sección 11.1 hemos visto que para soluciones estáticas y con poca curvatura la componente g_{tt} de la métrica es proporcional al potencial gravitatorio newtoniano. Por lo tanto comparando (11.13) y g_{tt} de (12.11), vemos que efectivamente podemos identificar el potencial newtoniano $\Phi = -M/r$. Dado que en la mecánica newtoniana el potencial gravitatorio de un objeto esférico con masa m viene dado por $\Phi = -G_N m/r$, está claro que la solución de Schwarzschild describe un objeto esférico y estático en el origen y además podemos interpretar la constante de integración M como

$$M = G_N m. \tag{12.12}$$

En otra palabras, M es una medida para la masa del objeto que causa la curvatura del espacio. A veces se denomina a M como la *masa geométrica* de la solución de Schwarzschild.

Aquí encontramos un primer ejemplo de como las constantes de integración que surgen de las ecuaciones de movimiento distinguen entre diferentes soluciones con las mismas condiciones de contorno. La ecuación de Einstein dice que el contenido de energía y materia no determina la geometría del espaciotiempo por completa, sino sólo la parte descrita por su tensor de Ricci. Por lo tanto, diferentes métricas, con tensores de Riemann distintas, pero con el mismo tensor de Ricci, son soluciones a la ecuacion de Einstein con el mismo tensor de energía-momento. Y una manera en que se diferencian estas métricas es en el valor de las constantes de integración, como vemos aquí: tanto el espacio de Minkowski, como la solución de Schwarzschild son soluciones de (12.1), pero Schwarzschild tiene un valor no-trivial para la constante de integración M que surge al resolver (12.9), mientras Minkowski se caracteriza precisamente por tener $M = 0$.

Llegado este punto es útil dedicar unas palabras a cómo compaginar el hecho de que por un lado la solución de Schwarzschild es una solución del vacío, en ausencia de masas y energía, y por otra lado su interpretación es la de un espaciotiempo con una masa m en el origen. La solución de Schwarzschild (12.11), por lo menos la parte con $r > R_0$, corresponde a la parte exterior del campo gravitatorio causado por un objeto esférico con masa m y radio $R_0 > 2M$ y por esto su nombre más correcto es la *solución exterior de Schwarzschild*. Claramente el exterior de ese objeto masivo es vacío y la solución correspondiente es una solución del vacío. Existe también otra solución (también encontrada por Karl Schwarzschild en 1916) que describe la parte interior $r < R_0$ donde se encuentra la estrella o el planeta, llamada la *solución interior de Schwarzschild*. El tensor de energía-momento de la solución interior es un fluido perfecto, mientras la solución exterior, como ya hemos visto, es una solución de vacío. Para el caso de un planeta o una estrella con radio $R_0 > 2M$, las dos soluciones enlazan suavemente en $r = R_0$. En la siguiente sección discutiremos lo que pasa si el radio R_0 del objeto en el centro es menor que $2M$.

Otra propiedad de la solución de Schwarzschild (12.11), es que para grandes valores de la coordenada radial r, el factor $M/r \ll 1$ y la métrica se aproxima cada vez más a Minkowski en coordenadas esféricas. En realidad, esto es de esperar a la luz de la interpretación de la solución como un objeto masivo en el origen. De la ley de Newton sabemos que la fuerza gravitatoria de un objeto masivo decae como $1/r^2$, es decir, a grandes distancias la influencia de la presencia del objeto es despreciable y el espacio se reduce a Minkowski. Soluciones que tienen esta propiedad se llaman *asintóticamente planas* y se las puede ver como objetos aislados.

Finalmente, existe un teorema de unicidad para la solución de Schwarzschild, llamado el *teorema de Birkhoff*. Este teorema dice que la solución exterior (12.11) es la única solución esféricamente simétrica de las ecuaciones del vacío. En particular, no existen soluciones de vacío con simetría esférica que no sean estáticas y si una solución de esta clase es estática, es Schwarzschild. En otras palabras una bola de materia puede contraerse o una estrella puede explotar conservando su simetría esférica, la solución exterior siempre será la misma métrica estática (12.11).

Aunque la demostración del teorema no es demasiado difícil, no entraremos en detalles, ya que nos llevaría fuera del alcance de estas notas. La demostración consiste básicamente en demostrar que el Ansatz más general con simetría esférica es el Ansatz (12.3), sustituyendo las funciones $A(r)$ y $B(r)$ por $A(t,r)$ y $B(t,r)$. Las ecuaciones de Einstein en seguida restringen la dependencia de A y B a ser funciones únicamente de r y el resto de la demostración es la misma que la que hemos hecho antes. El teorema de Birkhoff es un caso especial del teorema de "no-hair" (no tiene pelo) para agujeros negros en general, que discutiremos en más detalle en el Capitulo **??**. Este teorema dice que cualquier agujero negro clásico (no cuántico) está caracterizado por sólo tres cantidades físicas: la masa, la carga eléctrica y el momento angular. En el caso de soluciones esféricas del vacío, tanto la carga como el momento angular son cero y por lo tanto sólo el valor de M caracteriza la solución.

12.2. Estructura causal de la solución de Schwarzschild

Miramos ahora a la solución (12.11) entera, incluido la parte de $r < 2M$ para estudiar las singularidades. No es difícil ver que el elemento de linea (12.11)

$$ds^2 = \Big(1 - \frac{2M}{r}\Big)dt^2 - \Big(1 - \frac{2M}{r}\Big)^{-1}dr^2 - r^2\Big(d\theta^2 + \sin^2\theta d\varphi^2\Big), \qquad (12.13)$$

es degenerado o divergente para varios valores de las coordenadas. Un primer ejemplo es cuando $\theta = 0$ y $\theta = \pi$, porque entonces $g_{\varphi\varphi} = 0$. Sin embargo, esta degeneración es obviamente un artefacto del uso de las coordenadas esféricas. Cualquier rotación $SO(3)$ movería esta singularidad a otro sitio que antes era perfectamente regular y dejaría regular a los puntos conflictivos iniciales. Es más, un cambio a coordenadas cartesianas desharía por completo esta singularidad. La apa-

rente singularidad en $\theta = 0, \pi$ no es por lo tanto más que una singularidad de coordenadas, que no nos preocupará más.

Otras singularidades son $r = 0$ y $r = 2M$ donde g_{tt} es respectivamente singular y cero (y g_{rr} justo al revés), aunque aquí no está tan claro si son otra vez singularidades de coordenadas o si son físicas. Ya hemos dicho en la sección 10.4 que la manera de distinguir singularidades físicas de singularidades de coordenadas es mirando los *invariantes de curvatura*. Para nuestro caso de la solución de Schwarzschild (12.13), tanto R como $R_{\mu\nu}R^{\mu\nu}$ son de poca utilidad, ya que ambos son idénticamente cero por construcción, debido al hecho de que la solución de Schwarzschild es Ricci-plana. Sin embargo, calculando el llamado *invariante de Kretschmann* $R_{\mu\nu\rho\lambda}R^{\mu\nu\rho\lambda}$, obtenemos para la métrica (12.13) que

$$R_{\mu\nu\rho\lambda}R^{\mu\nu\rho\lambda} = \frac{48M^2}{r^6}, \tag{12.14}$$

es decir, el invariante de curvatura diverge para $r \to 0$, pero queda perfectamente regular para $r = 2M$. El punto $r = 0$ representa por lo tanto una singularidad física, mientras $r = 2M$ resulta ser una singularidad de coordenadas.[2]

Sin embargo, en contraste con la singularidad de coordenadas en $\theta = 0$ y $\theta = \pi$, la singularidad de coordenadas en $r = 2M$ sí tiene un significado físico. El radio $r = 2M$ se llama el *radio de Schwarzschild* y juega un papel importante en la física de la solución (12.13). Ahora, el hecho de que las llamadas *coordenadas de Schwarzschild* (t, r, θ, φ) sean singulares en $r = 2M$, hace que estas no sean válidas en el radio de Schwarzschild. Veremos que son las coordenadas adecuadas para describir la región donde $r > 2M$, pero para $r < 2M$ son poco fiables y para $r = 2M$ completamente inútiles.

Una de las maneras de ver que algo físicamente no-trivial pasa en el radio de Schwarzschild es mirando el efecto Doppler gravitatorio. En la sección 11.4 hemos visto que el corrimiento hacia el rojo de una señal emitida desde un punto con coordenada r_e y medida en un punto con coordenada r_d en una solución estática viene dada por (11.59)

$$T_d = T_e \sqrt{\frac{g_{tt}(r_d)}{g_{tt}(r_e)}} = T_e \sqrt{\frac{1 - \frac{2M}{r_d}}{1 - \frac{2M}{r_e}}}, \tag{12.15}$$

donde T_e y T_d son los periodos de la señal en el momento de emisión y detección respectivamente. Por lo tanto, cuando un reloj que cae hacia el centro emite pulsos con periodo T_e, un observador fijo en $r = r_d$ notará que el periodo T_d medido por él es tanto mayor cuanto más se aproxima el reloj al radio de Schwarzschild, hasta tal punto que el efecto Doppler es infinito cuando el reloj está en $r = 2M$. El radio de Schwarzschild es una *superficie de corrimiento infinito hacia el rojo*.

Hay otra indicación más, si miramos la estructura causal de la solución de Schwarzschild, a través de las geodésicas nulas radiales, las geodésicas nulas que van radialmente de $r = 0$ a $r = \infty$ ó al revés. En particular tienen $d\theta = d\varphi = 0$, de modo que su ecuación viene dada por (7.53)

$$\Big(1 - \frac{2M}{r}\Big)\dot{t}^2 - \Big(1 - \frac{2M}{r}\Big)^{-1}\dot{r}^2 = 0. \tag{12.16}$$

La expresión para t en función de r es por lo tanto,

$$\Big(\frac{dt}{dr}\Big)^2 = \Big(1 - \frac{2M}{r}\Big)^{-2}, \tag{12.17}$$

[2]Estrictamente hablando la regularidad de $R_{\mu\nu\rho\lambda}R^{\mu\nu\rho\lambda}$ en $r = 2M$ no es suficiente para concluir que se trata de una singularidad de coordenadas, ni tampoco el hecho de que numerosos otros invariantes sean finitos en este punto. El verdadero argumento es, que existe un cambio de coordenadas tal que $r = 2M$ se vuelve regular, como veremos dentro de poco.

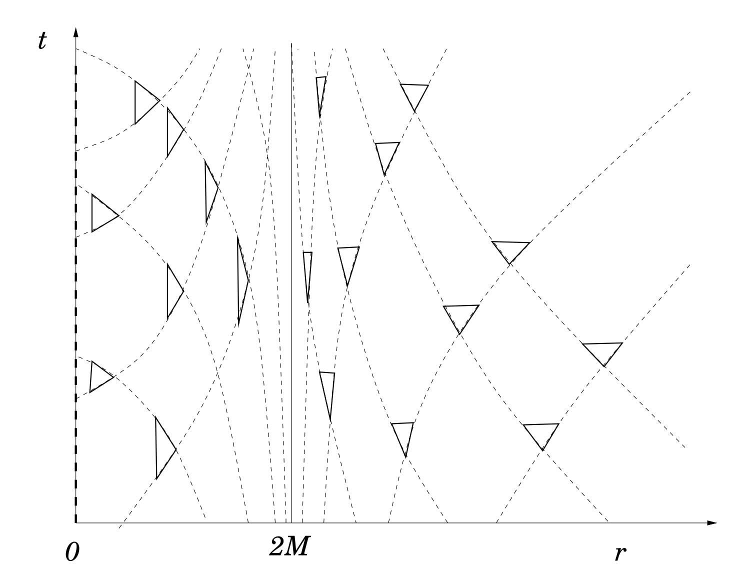


Figura 12.1: *Las geodésicas nulas radiales en la solución de Schwarzschild en las coordenadas de Schwarzschild* (t, r, θ, φ). *Para* $r \gg 2M$ *los conos de luz se comportan como en Minkowski, pero más cerca del radio de Schwarzschild los conos se cierran y degeneran en* $r = 2M$. *Dentro del radio de Schwarzschild los conos han cambiado de orientación y apuntan hacia la singularidad en* $r = 0$. *En estas coordenadas sin embargo parece que no es posible cruzar el radio de Schwarzschild desde el exterior.*

cuya solución viene dada por

$$t = \pm\Big(r + 2M \log|r - 2M| + C_0\Big), \tag{12.18}$$

donde C_0 es una constante de integración y los signos indican la dirección de las geodésicas: el signo posivito corresponde a geodésicas salientes y el signo negativo a geodésicas entrantes.

Lejos del centro, para $r \gg 2M$, el segundo término es despreciable frente al primero y los conos de luz se comportan más o menos como en Minkowski, es decir, hacen ángulos de 45° con los ejes t y r. Esto es de esperar, ya que la solución de Schwarzschild es asintóticamente plana y se aproxima a Minkowski para $r \to \infty$. Sin embargo, más cerca del radio de Schwarzschild, el término logarítmico se vuelve más importante, lo que indica que a los rayos de luz les cuesta más salir del pozo potencial creado por el campo gravitatorio. Concretamente, los conos de luz se cierran cada vez vez cuanto más nos acercamos al radio de Schwarzschild y en $r = 2M$ están completamente degenerados (vease Figura 12.1).[3] Dentro del radio de Schwarzschild, tanto las geodésicas entrantes como las salientes apuntan hacia la singularidad y los conos han cambiado de orientación y apuntan hacia la singularidad física en $r = 0$. Esto es una consecuencia del hecho de que para $r < 2M$, las componentes g_{tt} y g_{rr} de la métrica cambian de signo y las coordenadas t y r intercambian los papeles de coordenadas temporal y espacial. Veremos en breve el significado físico de esta propiedad.

En la Figura 12.1 parece que ninguna señal de luz es capaz de cruzar el radio de Schwarzschild, ni de fuera hacia dentro, ni al revés. Dado que una partícula masiva siempre se mueve dentro de su cono de luz, parece que tampoco ninguna partícula masiva, ni ningún observador es capaz de entrar a la zona $r < 2M$. En las coordenadas de Schwarzschild, una partícula que se cae hacia el centro, se parecer acercar asintóticamente al radio de Schwarzschild, pero nunca lo cruza.

[3]Esta es en realidad la causa de la singulardad de coordenadas en el radio de Schwarzschild: las direcciones nulas $x^{\pm} = t \pm (r + 2M \log|r - 2M|)$ son linealmente independientes en todos la variedad, salvo en $r = 2M$, de modo que allí formarían una base degenerada.

Pero esto resulta ser una conclusión errónea, debido a que las coordenadas (t, r, θ, φ) no son apropiadas para describir lo que pasa cerca del radio de Schwarzschild. En 1939, el físico estadounidense Robert Oppenheimer (1904 - 1967) se dió cuenta de que el tiempo de caída es muy diferente, si lo expresamos en el tiempo propio de la partícula en caída libre. Para esto es necesario hacer un estudio de las geodésicas radiales temporales.

Para calcular las geodésicas temporales radiales estamos interesados en la componente temporal de la ecuación (7.42),

$$\ddot{t} \; + \; 2\Gamma^t_{tr}\dot{t}\dot{r} \; = \; 0, \tag{12.19}$$

donde el punto se refiere a la derivación con respecto al tiempo propio τ. Para nuestro caso de la solución de Schwarzschild, podemos simplificar esta expresión, multiplicando la ecuación por g_{tt} y escribiendo el símbolo de Christoffel explícitamente en función de la métrica con la fórmula (8.7), de modo que (12.19) se convierte en (ejerc.)

$$\frac{d}{d\tau}\Big(g_{tt}\,\dot{t}\Big) \; = \; 0, \tag{12.20}$$

lo que se puede integrar directamente a

$$\Big(1 - \frac{2M}{r}\Big)\,\dot{t} \; = \; 1. \tag{12.21}$$

La constante de integración 1 ha sido elegida tal que la ecuación corresponde a la de una partícula que cae desde infinito con velocidad inicial cero, puesto que de (12.21) deducimos que $d\tau \approx dt$ para $r \gg 2M$. En otras palabras, cerca del infinito los efectos de dilatación temporal especial-relativista entre el observador en caída libre y el observador midiendo con el tiempo t son despreciables. Por lo tanto podemos concluir que la coordenada t corresponde al tiempo propio de un observador asintótico en el infinito, lejos del centro de la solución. Esto ya es una primera indicación de que el sistema de coordenadas (t, r, θ, φ) no es muy apropiado cerca del radio de Schwarzschild, ya que sabemos que el tiempo no corre igual arriba y abajo en un potencial gravitatorio.

Por lo tanto, las ecuaciones para las geodésicas temporales vienen dadas por (12.21), junto con la ecuación (7.53), que en nuestro caso es de la forma

$$\Big(1 - \frac{2M}{r}\Big)\,\dot{t}^2 \; - \; \Big(1 - \frac{2M}{r}\Big)^{-1}\dot{r}^2 \; = \; 1. \tag{12.22}$$

Esta última expresión la podemos rescribir como

$$\dot{r}^2 \; = \; \Big(1 - \frac{2M}{r}\Big)^2\dot{t}^2 \; - \; \Big(1 - \frac{2M}{r}\Big) \; = \; \frac{2M}{r}, \tag{12.23}$$

donde en la última igualdad hemos utilizado la ecuación (12.21). Lo más sencillo es ver (12.23) como una ecuación diferencial para τ en función de r,

$$\frac{d\tau}{dr} = -\sqrt{\frac{r}{2M}}, \tag{12.24}$$

con solución (ejerc.)

$$\tau \; = \; \frac{2}{3\sqrt{2M}}\Big(r_0^{3/2} - r^{3/2}\Big). \tag{12.25}$$

En otras palabras, cayendo desde una distancia r_0, un observador $\mathcal{O}'$ en caída libre llega al radio de Schwarzschild $r = 2M$ e incluso a la singularidad $r = 0$ en un intervalo de tiempo propio τ finito. Por lo tanto, aunque no lo parece de la Figura 12.1, es perfectamente posible llegar al radio de Schwarzschild, cruzarlo y llegar a la singularidad. El problema está en que la Figura 12.1 está dibujado en coordenadas (t, r, θ, φ), que son las coordenadas de un observador $\mathcal{O}$ en $r = \infty$.

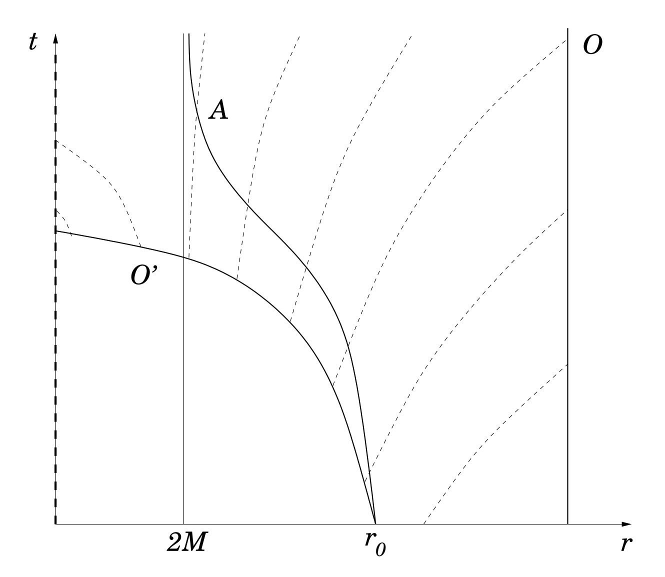


Figura 12.2: *Trayectoria de un observador $\mathcal{O}'$ en caída libre visto por un observador lejano $\mathcal{O}$: un observador en caída libre $\mathcal{O}'$ alcanza la singularidad después de un tiempo propio finito. Los señales de luz emitido por el observador en caída libre (líneas puntuales) siguen geodésicas radiales nulas y tardan cada vez más en llegar al observador lejano $\mathcal{O}$, que cree erróneamente que $\mathcal{O}'$ sigue la trayectoria A y que nunca cruza el radio de Schwarzschild. Una vez pasado el radio de Schwarzschild $\mathcal{O}'$ ya no tiene manera de comunicarse con $\mathcal{O}$ para convencerle de lo contrario.*

La Figura 12.1 refleja por lo tanto no lo que realmente ocurre con $\mathcal{O}'$ cerca del radio de Schwarzschild, sino más bien lo que pasa con el observador $\mathcal{O}'$ según el punto de vista del observador $\mathcal{O}$. Y lo único que tiene éste para enterarse son las señales que emite $\mathcal{O}'$ periódicamente. Pero ya hemos visto que estas señales siguen las geodésicas nulas radiales (12.18) y llegan por lo tanto a $\mathcal{O}$ con intervalos cada vez mayores, como consecuencia de sufrir un corrimiento hacia el rojo cada vez mayor. $\mathcal{O}$ ve a $\mathcal{O}'$ acercarse asintóticamente al radio de Schwarzschild, pero nunca lo ve cruzarlo, ya que la señal emitida por $\mathcal{O}'$ en $r = 2M$ llega a $\mathcal{O}$ en el momento $t = \infty$. Las señales que $\mathcal{O}'$ emite después de cruzar el radio de Schwarzschild no salen hacia el exterior y no llegarán nunca a $\mathcal{O}$, ni siquiera después de una cantidad infinita de tiempo t (véase Figura 12.2).

12.3. Las coordenadas de Eddington-Finkelstein

Ya hemos visto algunas características de la física que ocurre cerca del radio de Schwarzschild e incluso hemos deducido algunas propiedades de geodésicas temporales y nulas para $r < 2M$, aunque siempre en el sistema de coordenadas (t, r, θ, φ), que no es muy fiable cerca del radio de Schwarzschild. Para entender cómo conectan las regiones dentro y fuera del radio de Schwarzschild y para entender mejor la estructura causal de la solución, es necesario utilizar unas coordenadas que no sean singulares en el radio de Schwarzschild y que describan bien el comportamiento de objetos físicos (masivos o sin masa) en $r = 2M$.

Insprirados por la expresión (12.18) para las geodesicas radiales nulas entrantes (con signo menos), consideramos la siguiente redefinición de la coordenada temporal

$$\tilde{t} \;=\; t \;+\; 2M \log(r - 2M). \tag{12.26}$$

Nótese que el cambio de coordenadas es tal que lejos del radio de Schwarzschild, el término logarítmico es despreciable con repecto al término lineal y $\tilde{t} \approx t$. Sin embargo cuando $r \sim 2M$, la

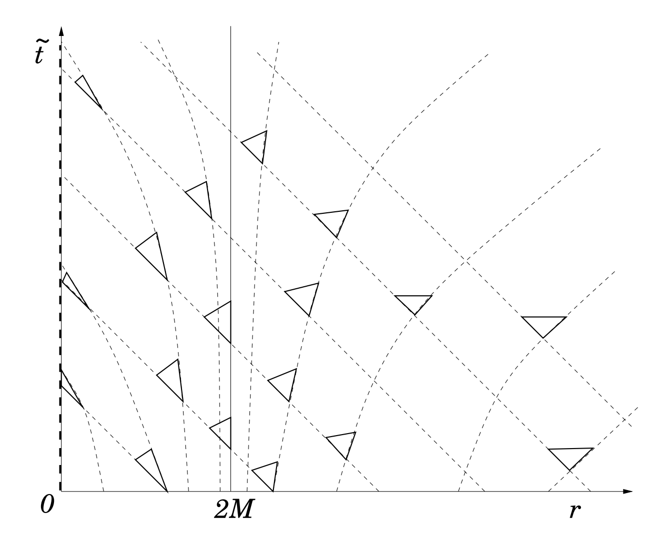


Figura 12.3: *La solución de Schwarzschild en coordenadas avanzadas de Eddington-Finkelstein. Los conos de luz se inclinan tanto más cuanto más cerca de la singularidad. El radio de Schwarzschild $r = 2M$ es especial, porque allí la inclinación es tanto que ni las geodésicas nulas salientes llegan al infinito. El radio de Schwarzschild forma un horizonte deja entrar influencias causales, pero no salir. Un observador exterior no puede ver lo que ocurre dentro y un observador dentro no tiene manera de evitar la singularidad.*

coordenada temporal $\tilde{t}$ difiere mucho del tiempo t, medido por el observador lejano $\mathcal{O}$.

No es difícil ver que las expresiones (12.18) para las geodésicas radiales nulas en coordenadas de Eddington-Finkelstein son

$$\tilde{t} \;=\; -r + C_0, \qquad\qquad \tilde{t} \;=\; r \;+\; 4M\log|r-2M| + C_0, \tag{12.27}$$

para las geodésicas entrantes y salientes respectivamente. Nótese que las geodésicas entrantes tienen una forma particularmente fácil (esta es la razón por la que hemos cambiado a estas coordenadas) y que no divergen en $r = 2M$, sino que pasan desde $r = \infty$ hasta $r = 0$ sin problema. En particular, veremos que estas coordenadas $(\tilde{t}, r, \theta, \varphi)$ no son singulares en el radio de Schwarzschild y nos permiten trazar sin problemas trayectorias hasta $r < 2M$. Las coordenadas $(\tilde{t}, r, \theta, \varphi)$ se llaman las *coordenadas (avanzadas)*[4] *de Eddington-Finkelstein*, introducidas por David Finkelstein (1929 -) en 1958.[5]

En la Figura 12.3 se puede ver que el comportamiento de los conos de luz en las coordenadas de Eddington-Finkelstein es distinto que en las coordenadas de Schwarzschild. Para $r \gg 2M$, los conos se comportan como en el espacio de Minkowski, igual que en las coordenadas de Schwarzschild, pero cerca del $r = 2M$, los conos empiezan a inclinarse hacia la singularidad. Las geodésicas radiales nulas entrantes siempre mantienen el mismo ángulo de 45° con el eje r, pero las geodésicas salientes tienen un ángulo cada vez mayor, cuanto más se acercan a $r = 2M$, ya que a las geodésicas salientes les cuesta cada vez más salir hacia el infinito, debido a la curvatura del espacio. El radio de Schwarzschild es un punto especial, porque allí la inclinación de los conos

[4]Se suele añadir a estas coordenadas el adjetivo "avanzada", porque también se pueden construir, a partir de las geodésicas salientes, unas coordenadas $(\bar{t}, r, \theta, \varphi)$, con $\bar{t} \;=\; t \;-\; 2M\log(r-2M)$, llamadas las *coordenadas retardadas de Eddington-Finkelstein*. Estará claro que la diferencia entre las coordenadas avanzadas y retardadas consiste básicamente en el intercambio de las propiedades de las geodésicas entrantes y salientes. Aquí supondremos que si no añadimos el adjetivo explícitamente, nos referiremos a las avanzadas. Discutiremos brevemente las retardadas un poco más adelante.

[5]El primero en construir un sistema de coordenadas que no fuese singular en $r = 2M$ fue Arthur Eddington, aunque no parece haberse dado cuenta del significado físico. El primero en entender que la singularidad en el radio de Schwarzschild es una singularidad de coordenadas y no una singularidad física fue el cura y físico belga George Lemaître (1894-1966) en 1933.

de luz es tanta que las geodésicas nulas salientes forman un ángulo de 90° con el eje r. Una señal de luz, emitido desde allí radialmente hacia el exterior, se quedará a distancia fija $r = 2M$ del centro. Para $r < 2M$, la inclinación de los conos es aún mayor e incluso las geodésicas salientes están dirigidas hacia el centro y acaban en la singularidad.

De estas observaciones podemos derivar dos efectos físicos importantes de la solución de Schwarzschild. Primero, dado que tanto las señales de luz, tanto entrantes como salientes, emitidas desde la región $r < 2M$ están dirigida hacia el centro, vemos que el radio de Schwarzschild actúa como un *horizonte de sucesos*. Como escribió el ísico estadounidense David Finkelstein (1929 -) en 1958, *"la superficie $r = 2M$ es [...] verdaderamente una membrana unidireccional: las influencias causales sólo pueden atravesarla en una dirección."* Ningún evento en la región $r < 2M$ puede influenciar lo que ocurre fuera del radio de Schwarzschild ya que los conos de luz no alcanzan hasta la región exterior. Para un observador exterior es por lo tanto imposible obtener información sobre lo que occure dentro, ya que desde allí ninguna señal puede escapar hacia el infinito. Lo mismo ocurre para partículas (u observadores) masivos: una vez cruzado el horizonte, ya no hay manera de volver a la región asintóticamente plana.

Una singularidad rodeada por un horizonte de sucesos, que evita ver lo que pasa dentro del horizonte, es un *agujero negro*, uno de los fenómenos más violentos en la Naturaleza. Obsérvese que el horizonte $r = 2M$ no es una superficie temporal, sino una superficie nula, ya que es la superficie en la que se mueven las señales de luz emitidas hacia el exterior desde este punto. Merece la pena enfatizar que la apareción del horizonte es una propiedad global, debido a la orientación de los conos de luz en toda la región $r \leq 2M$, es decir, debido a la estructura causal de la solución de Schwarzschild. Localmente es espacio es completamente regular en $r = 2M$, sin ninguna propiedad geométrica particular. Un observador puntual en caída libre no notaría ningún efecto físico al cruzar el horizonte. Esto es una consecuencia del Principio de Equivalencia, que sigue válido en cualquier punto del espaciotiempo en general y el el radio de Schwarzschild en particular.

El segundo efecto físico se trata de la singularidad: dado que tanto las geodésicas entrantes como las salientes acaban en la singularidad y dado que cualquier partícula (u observador) masivo siempre queda confinado dentro de su propio cono de luz, está claro que también las partículas masivas acabarán inevitablemente en la singularidad. Es imposible quedarse en reposo a un $r = r_0 < 2M$ fijo del centro. Cualquier curva temporal o nula que cruce el horizonte está destinada a darse con la singularidad y desaparecer del espaciotiempo. Esto es una consecuencia de que para $r < 2M$ la coordenada r es una coordenada temporal, como ya mencionamos antes, y que la singularidad en $r = 0$ es una superficie espacial (no temporal) y además está en el futuro causal de cualquier observador que se atreve cruzar el horizonte. Una vez cruzado $r = 2M$, avanzar en la dirección radial y acabar en la singularidad es igual de inevitable que en el espacio de Minkowski evolucionar en el tiempo y cruzar una superficie temporal $t = t_0$.

En realidad ya hemos visto algunas de estas propiedades de la geometría dentro del radio de Schwarzschild de una u otra forma en las Figuras 12.1 y 12.2, pero allí aparecieron en las coordenadas (t, r, θ, φ), que son solamente fiables en la región $r > 2M$. Ahora, en las coordenadas de Eddington-Finkelstein podemos afirmar estas conclusiones, ya que son regulares en el horizonte, como se puede ver de la forma de la métrica en estas coordenadas. De (12.26) vemos que

$$dt \;=\; d\tilde{t} \;-\; \frac{2M}{r}\Big(1 - \frac{2M}{r}\Big)^{-1} dr, \qquad (12.28)$$

de modo que sustituyendo esta expresión en la métrica (12.13) obtenemos la siguiente expresión para la solución de Schwarzschild en coordenadas avanzadas de Eddington-Finkelstein:

$$ds^2 \;=\; \Big(1 - \frac{2M}{r}\Big)d\tilde{t}^2 \;-\; \frac{4M}{r}d\tilde{t}dr \;-\; \Big(1 + \frac{2M}{r}\Big)dr^2 \;-\; r^2 d\Omega_2^2. \qquad (12.29)$$

Esta claro que la métrica ya no diverge en $r = 2M$, aunque g_{tt} sea cero.[6]

[6]Estrictamente hablando el cambio de coordenadas (12.26) sólo está definido para $r > M$. Sin embargo es fácil ver

Sin embargo, obsérvese que, debido al término cruzado $g_{\tilde{t}r}$, en estas coordenadas la métrica (12.29) ya no tiene la simetría $\tilde{t} \to -\tilde{t}$ que uno esperaría de una solución estática. Pero ya hemos dicho antes que, en lugar de las coordenadas avanzadas de Eddington-Finkelstein (12.26), podíamos haber elegido las coordenadas retardadas

$$\bar{t} = t - 2M\log(r - 2M), \tag{12.30}$$

de modo que la métrica sería

$$ds^2 = \Big(1 - \frac{2M}{r}\Big)d\bar{t}^2 + \frac{4M}{r}d\bar{t}dr - \Big(1 + \frac{2M}{r}\Big)dr^2 - r^2 d\Omega_2^2. \tag{12.31}$$

Obsérvese que la diferencia entre estas dos expresiones es el signo del término cruzado, de modo que la transformación $\tilde{t} \rightleftarrows \bar{t}$ mapea una métrica en la otra. En la región $r > 2M$, las transformaciones (12.26) y (12.30) mapean la métrica (12.13) en (12.29) y (12.31) respectivamente, de modo que fuera del radio de Schwarzschild, estas tres métricas corresponden a la misma solución. Sin embargo, dentro del radio de Schwarzschild (12.29) y (12.31) no son equivalentes, ya que el cambio de coordenadas $\tilde{t} \rightleftarrows \bar{t}$ no está bien definido para $r < 2M$. El hecho de que estas dos métricas no sean físicamente equivalentes se ve en el comportamiento de las geodésicas radiales nulas. No es difícil ver que en las coordenadas $(\bar{t}, r, \theta, \varphi)$ son las geodésicas salientes las que siempre tienen un ángulo constante de 45° con el eje r y que los conos de luz cerca de la singularidad se inclinan hacia el exterior (ejerc.). El radio de Schwarzschild $r = 2M$ en (12.31) corresponde a una membrana que sólo deja pasar influencias causales desde el interior hacia el exterior, un tipo de "agujero blanco", justo al revés que en las coordenadas $(\tilde{t}, r, \theta, \varphi)$.

La solución de esta extraña paradoja es que las coordenadas avanzadas y retardadas de Eddington-Finkelstein describen la misma región del espaciotiempo para $r > 2M$, pero la parte con $r < 2M$ en coordenadas $(\tilde{t}, r, \theta, \varphi)$ no es la misma que la parte con $r < 2M$ en coordenadas $(\bar{t}, r, \theta, \varphi)$. Las coordenadas avanzadas describen una región asintóticamente plana con una singularidad cubierta por un horizonte en el futuro. Una vez cruzado este horizonte es imposible volver a la parte asintóticamente plana. Las coordenadas retardadas describen la misma región asintóticamente plana, pero con una singularidad cubierta por un horizonte en el pasado. Influencias causales pueden salir de la singularidad, cruzar este horizonte y llegar a la región asintóticamente plana, pero una vez cruzado este horizonte es imposible volver a la singularidad original (véase Figura 12.4).

Matemáticamente lo que ocurre es que ninguna de las dos coordenadas describe la variedad completa, sólo parches de ella. En la Figura 12.4, la parte asintóticamente plana corresponde a la región I y el horizonte $r = 2M$ corresponde con las líneas continuas de 45°. Detrás del horizonte hay dos regiones, II y II', mutuamente simétricas en el tiempo, cada una con su propia singularidad. Las coordenadas avanzadas cubren las regiones I y II, mientras las coordenadas retardadas las regiones I y II'.

Se podría preguntar si existe un sistema de coordenadas que cubra todas estas regiones a la vez. La respuesa es sí, las llamadas *coordenadas de Kruskal*, aunque nos llevaría demasiado lejos dar la derivación en detalle. En estas coordenadas la métrica viene dada por

$$ds^2 = \frac{16M^2}{r}e^{-r/2M}\Big(dT^2 - dR^2\Big) - r^2 d\Omega_2^2, \tag{12.32}$$

donde las coordenadas T y R están relacionadas con las coordenadas de Schwarzschild t y r a través de

$$T = e^{r/4M}\sqrt{r - 2M}\ \sinh\Big(\frac{t}{4M}\Big), \qquad R = e^{r/4M}\sqrt{r - 2M}\ \cosh\Big(\frac{t}{4M}\Big), \tag{12.33}$$

que el cambio de coordenadas $\tilde{t} = t + 2M\log(2M - r)$ aplicado a la parte $r < 2M$ de la métrica (12.13) lleva a la misma expresión, lo que demuestra que las dos partes disconexas $r < 2M$ y $r > 2M$ de (12.13) en realidad describen la misma solución, y que las coordenadas de Eddington-Finkelstein $(\tilde{t}, r, \theta, \varphi)$ cubren ambas regiones.

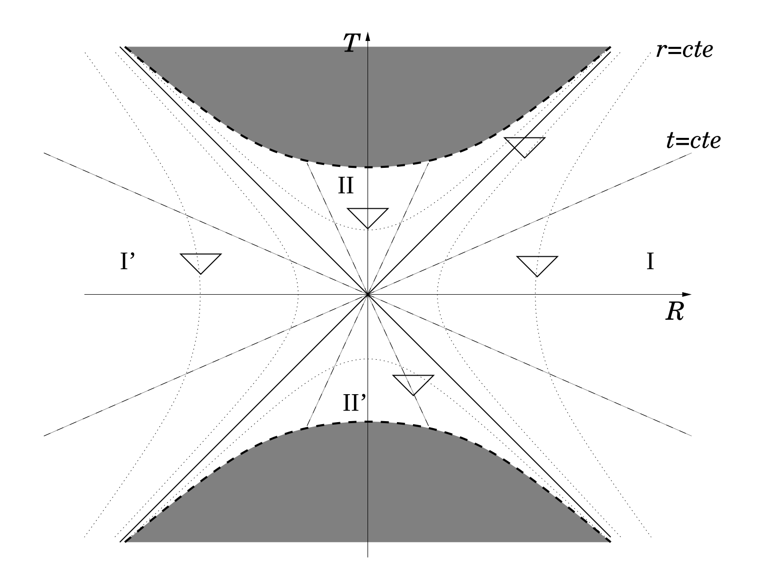


Figura 12.4: *La solución de Schwarzschild en coordenadas de Kruskal: la luz se mueve en líneas rectas con ángulos de 45°, de modo que la orientación de los cono de luz es la misma en todo el espacio. La solución de Schwarzschild consiste de 2 partes asintóticamente planas I y I′, conectadas por un agujero de gusano en $R = T = 0$. Las regiones I y I′ están separado por un horizonte de sucesos de las regiones II y II′, que contienen una singularidad espacial en el futuro y el pasado respectivamente. Las singularidades estan representadas por la línea gorda interrumpida y los horizontes por la línea gorda ininterrumpida. Líneas de t=cte y r=cte estan representadas respectivamente por las líneas finas interrumpidas y las líneas de puntos.*

de modo que

$$T^2 - R^2 = e^{r/2M}(r - 2M) \tag{12.34}$$

No sólo estas coordenadas cubren las regiones I, II y II′ a la vez, sino también resulta que la variedad tiene otra región asintóticamente plana extra, la región I′, que ninguna de las coordenadas de Eddington-Finkelstein podía ver. Se puede demostrar que las coordenadas de Kruskal forman la *extensión máxima* de la solución de Schwarzschild, es decir, que cubren toda la variedad, sin que queden más regiones por descubrir.

En estas coordenadas, tanto las geodésicas entrantes como las salientes forman ángulos de 45° con los ejes T y R, de modo que los conos de luz (los triángulos sólidos en Figura 12.4) se comportan como en el diagrama del espacio de Minkowski. Sin embargo, de (12.34) vemos que las curvas con $r =$cte forman hipérbolas y los planos $t =$cte forman ángulos más grandes para tiempos más avanzados. El horizonte $r = 2M$ es una superficie nula, ya que también forma un ángulo de 45° con los ejes y se ve claramente que una partícula que entra en la región II no podrá emitir señales a la región I y acabará en la singularidad (la línea gorda punteada), que es una superficie espacial y por lo tanto inevitable.

En la región II′ pasa justamente lo contrario: cualquier señal emitida desde aquí forzosamente tiene que salir hacia las regiones asintóticamente planas I o I′. Una vez en estas regiones, ya no hay manera de volver a la región II′. Ningún evento en las regiones I o I′ pueden influenciar lo que ocurre en II′, ya que la región II′ está en el pasado de I y I′. Nótese que tanto en la zona II como en la zona II′, la coordenada t es espacial y r es temporal, como ya habíamos mencionado antes.

En las coordenadas T y R, la solución no sólo es completamente regular en $r = 2M$, sino que también es simétrica bajo la inversión de la coordenada temporal $T \to -T$, como uno espera de una solución estática. Obsérvese que bajo $T \to -T$ las regiones I y I′ son invariantes y II y II′

se intercambian. Pero más sorprendente es que también hay una simetría bajo inversión de la coordenada radial $R \to -R$, que deja II y II′ invariantes e intercambia I y I′. La región I′ es por lo tanto un tipo de imagen de espejo de la región I, o sea, otro espacio asintóticamente plano, donde observadores y partículas pueden caer en la singularidad de la región II. Las regiones I y I′ están conectadas en el punto $T = R = 0$ por un agujero de gusano, llamado un *puente de Einstein-Rosen*, pero la estructura causal es tal que no pueden pasar influencias causales de una región a la otra. Observadores de una región sí pueden ponerse en contacto con observadores de la otra región, si los dos se atreven a entrar en la región II: una vez allí se podrán influenciar mutuamente antes de acabar en la singularidad, pero nunca podrán volver a influenciar eventos en sus antiguos paraderos I y I′.

12.4. Colapso gravitacional y formación de agujeros negros

Una se podría preguntar si agujeros negros tipo Schwarzschild realmente existen en la Naturaleza, o si sólo son una solución matemática, sin realidad física. La respuesta es un poco ambivalente: realmente se pueden llegar a formar agujeros negros, por ejemplo al final de la vida de estrellas muy masivas, pero a pesar de que tienen muchas de las características que hemos comentado en la sección anterior, no son exactamente como la extensión máxima de la solución de Schwarzschild. La gran diferencia está en que los agujeros negros en la Naturaleza están formados dinámicamente en un proceso de colapso gravitacional y por lo tanto no tienen la simetría de inversión temporal de una solución estática como la extensión maximal (12.32).

Ya hemos visto en la sección 12.1, que la solución (12.11) en realidad se corresponde con la parte exterior de un campo gravitatorio causado por un objeto con masa $m = M/G_N$ en el centro. En circunstancias normales, la masa ocupa una esfera con un radio R_0 mayor que $2M$, de modo que la solución exterior es la métrica (12.11) y la interior es la solución interior de Schwarzschild, que mencionamos antes. Sin embargo, si comprimimos la masa en un volumen más pequeño, la gravedad en la superficie aumentará, ya que el potencial gravitatorio varía como $\Phi = -G_N m/r$.[7]

Al comprimir la masa en un volumen más pequeño, aumentará también la velocidad de escape, la velocidad inicial necesario para que una partícula pueda salir del pozo potencial de un objeto masivo y llegar al infinito. Desde la ley de conservación de energía de la mecánica newtoniana, se puede calcular fácilmente que la velocidad de escape v_e de un objeto con masa m y radio R viene dada por

$$v_e = \sqrt{\frac{2G_N m}{R}}. \tag{12.35}$$

En 1795 el matemático francés Pierre Simon Laplace (1749 - 1827) se dió cuenta de que la velocidad de escape superaría la velocidad de la luz si se comprimiese toda la masa en un radio $R = 2G_N m$. Escribe en su *Traité de la Méchanique Céleste* en 1799: *Una estrella luminosa de la misma densidad que la tierra, y cuyo diametro es 250 veces mayor que el sol, no dejaría llegar por su atracción ningún rayo hasta nosotros; por lo tanto es posible que los más grandes cuerpos luminosos del Universo sean, por esa razón, invisibles para nosotros.* Sorprendentemente, el radio crítico para la velocidad de escape, calculado con métodos puramente newtonianos, coincide exactamente con el radio de Schwarzschild, el radio desde donde la luz ya no puede salir hacia el exterior.[8] La interpretación, sin embargo es distinta, ya que en la mecánica newtoniana, la velocidad de la luz no es un límite superior, de modo que la "estrella negra" de Laplace no es un agujero negro en el sentido estricto de la palabra.

[7] Esto no implica que la gravedad a distiancia $r_0 > R_0$ aumente: el potencial gravitatorio a distancia r_0 fuera de la masa es independiente del volumen que ocupa la masa. Si el Sol colapsara en un agujero negro, la trayectoria de la Tierra no cambiaría en absoluto.

[8] Quizá no es tan sorprendente: $G_N m$ es la única combinación con dimensión de longitud que se puede construir, y el factor 2 viene justamente de ajustar en (12.12) la constante de integración M con la mecánica newtoniana.

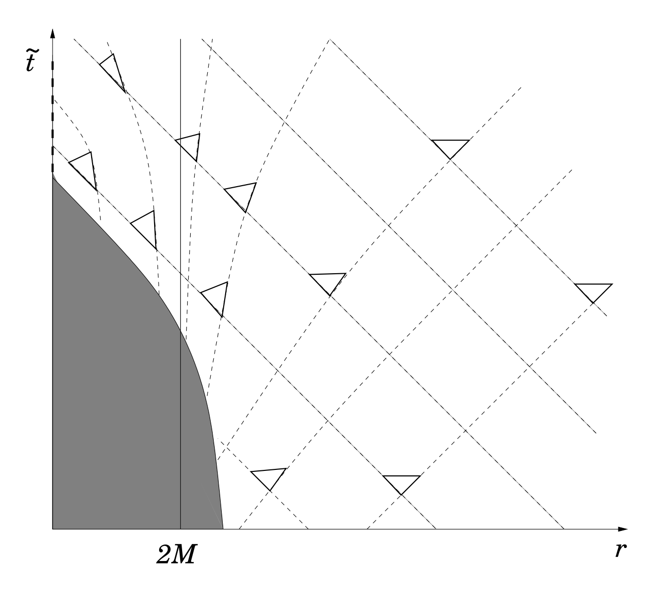


Figura 12.5: *El colapso gravitatorio en las coordenadas de Eddington-Finkelstein: una masa m ocupa inicialmente un volumen mayor que su radio de Schwarzschild $2G_Nm$, pero según se va contrayendo, aumenta la curvatura del espacio cerca del centro. Cuando la masa se haya comprimido dentro del radio de Schwarzschild, se forma un agujero negro, con su singularidad y su horizonte. Toda la materia de la masa original termina en la singularidad y desaparece del espaciotiempo.*

En la relatividad general, más que la velocidad de escape, la cantidad física importante es la curvatura del espaciotiempo: cuanto más se comprime la masa, tanto más aumenta la curvatura alrededor del objeto y una vez que toda la masa está comprimido en un volumen más pequeño que el radio de Schwarzschild, ya no hay manera de parar el colapso gravitacional. La curvatura es tanto que la luz se queda atrapada, ya que incluso las geodésicas nulas están dirigidas hacia el centro. Se forma por lo tanto un horizonte de sucesos y, debido al teorema de Hawking y Penrose, también una singularidad. Por la estructura causal del espaciotiempo dentro del horizonte, toda la materia del objeto original acabará en la singularidad y desparecerá del espaciotiempo (véase Figura 12.5), tal como se dió cuenta Oppenheimer en 1939.

Por lo tanto no es preciso disponer de una masa grande para poder formar un agujero negro, por lo menos, en principio. Más que de la masa, la formación de una agujero negro depende de la densidad: cualquier masa m puede formar un horizonte y una singularidad si se comprime dentro del radio de Schwarzschild correspondiente a esa masa, es decir en un volumen $r = 2G_Nm$. El radio de Schwarzschild del Sol es aproximadamente 3 km y el de la Tierra unos 9 mm.

En la práctica no hay fuerza en la Naturaleza capaz de comprimir ni el Sol, ni la Tierra dentro de sus respectivos radios de Schwarzschild. En el caso de planetas como la Tierra, la repulsión entre los electrones de los átomos es suficiente para contrarrestar la fuerza gravitatoria y prevenir un colapso gravitacional. En objetos más masivos, como estrellas, la presión gravitacional hacia dentro es tan grande que la materia forma un plasma tan caliente que hay fisión nuclear. En grandes líneas, 4 protones se juntan para formar un núcleo de helio. La energía térmica producida por estas reacciones nucleares contraresta la presión gravitatoria, de modo que la estrella se encuentra en un equilibrio térmico-gravitatorio.

Sin embargo, cuando a la estrella se le acaba el combustible, ya no es capaz de producir la energía términa necesaria para mantener el equilibrio. Lo que pasa entonces, depende básicamente de la masa de la estrella considerada. Para estrellas pequeñas y medianas, como el Sol, la gravedad comprimirá la estrella en un volumen comparable con la Tierra, con una densidad entre 10^4 hasta 10^9 kg/cm^3. Allí la presión del gas degenerado de electrones y átomos completamente

ionizados será lo suficiente para volver a mantener el equilibrio. La estrella se ha convertido en una *enana blanca*, llamada así por su tamaño y su color, debido a su alta temperatura.

En 1931 el astrofísico Subrahmanyan Chandrasekhar (1910 - 1995) demostró que una enana blanca no puede tener una masa mayor que unos 1,4 masas solares. Para estrellas con masas superiores al *límite de Chandrasekhar*, la presión del gas degenerado no es capaz de contrarrestar la fuerza gravitacional. En este caso, la estrella será comprimida en un radio de unos pocos kilómetros. La presión gravitatoria es tan fuerte que los electrones están comprimidos dentro de los núcleos de los átomos y reaccionarán con los protones para formar una bola inmensa de neutrones, llamada una *estrella de neutrones*. A estas alturas, la única fuerza que puede resistir la presión gravitacional es la repulsión fermiónica entre los neutrones. La densidad típica de una estrella de neutrones es de unos $10^{17}\ kg/cm^3$.

En 1939 Oppenheimer y collaboradores calcularon el equivalente del límite de Chandrasekhar para estrellas de neutrones, el llamado *límite de Oppenheimer-Volkov* y encontraron que la repulsión entre los neutrones no es suficiente para contrarrestar la fuerza gravitatoria en estrellas de neutrones con masas superiores a unos 2 masas solares. En este caso ya no hay ninguna fuerza capaz de controlar la compresión gravitacional y la estrella colapsará a su radio de Schwarzschild y más allá ...

Einstein y Eddington, el padre de la teoría de la evolución estelar, se opusieron firmemente a la idea de que un estrella podría colapsar a un solo punto y renunciaron a dar a la solución de Schwarzschild algún significado físico. Pero hoy en día sabemos que existen estrellas con masas de decenas y cientos de veces la del Sol, así que, a pesar de que puedan perder mucha masa a lo largo de su evolución, es posible que algunas estrellas saturen el límite de Oppenheimer-Volkov y se conviertan en agujeros negros.

Además, la evolución estelar no es la única fuente de agujeros negros. Los astrónomos lo sospecharon desde hace tiempo, pero en los últimos años han encontrado pruebas convincentes de que existen *agujeros negros supermasivos* en los centros de galaxias, con masas del orden de 10^9 masas solares y radios de Schwarzschild del orden de $10^9\ km$. Al otro lado de la escala se sospecha que también existen *agujeros negros primordiales*, que no se formaron en colapsos gravitacionales de estrellas, sino en las primeras fases de la existencia del universo, debido a las fluctuaciones de densidad. Se calcula que estas agujeros negros primordiales tendrían una masa del orden de $10^{12}\ kg$ (en comparación, la Tierra tiene una masa del orden de $10^{24}\ kg$) y un radio de Schwarzschild de unos $10^{-14}m$.

Queda por preguntar ¿qué es lo que pasa con un observador que se cae en un agujero negro? Ya sabemos que acabará inevitablemente en la singularidad, pero antes de esto, ¿qué ve y qué siente? Ya hemos dicho que el horizonte geométricamente hablando es un punto perfectamente regular y que un observador puntual en caída libre no notaría nada al cruzarlo. Por la estructura causal del espaciotiempo está claro que el observador no puede ver la singularidad mientras que está fuera del radio de Schwarzschild, ya que ninguna señal puede salir del horizonte para advertirle de lo que le espera. Pero tampoco una vez dentro es capaz de ver la singularidad: de los diagramas de espaciotiempo está claro que no pueden salir señales de la singularidad hacia un $r < 2M$. La coordenada radial es una coordenada temporal para $r < 2M$, así que querer ver la singularidad una vez dentro del radio de Schwarzschild es como querer ver en el futuro. El observador por lo tanto no ve el rostro de la singularidad hasta que da con ella.

Esto no implica que el viaje hacia $r = 0$ sea agradable, por lo menos no para observadores reales, es decir, observadores no puntuales. Objetos y observadores con una extensión espacial sufrirán grandes fuerzas de marea, debido a la inhomogeneidad del campo gravitatorio. En un potencial gravitatorio que varía como $1/r$, la fuerza de marea, es decir la diferencia entre la fuerza a distancia r y $r + \Delta r$ es proporcional a M/r^3, de modo que cerca de la singularidad, las fuerzas de marea son inmensas. No sólo la diferencia de la fuerza gravitacional entre la cabeza y los pies estira y desgarra al observador (o lo que queda de él) longitudinalmente como un potro de

tortura cósmico, también transversalmente le comprime en una superficie $r^2\Delta\Omega$ cada vez menor.

Estas fuerzas de marea son muy grandes, cerca de la singularidad, pero esto no necesariamente implica que lo sean en el horizonte. Para agujeros negros estelares y primordiales efectivamente lo son: para un agujero negro con la masa del Sol, un humano ya no sobreviviría las fuerzas a una distancia de 200 km, unos 66 radios de Schwarzschild. Pero en un agujero negro supermasivo de 10^9 masas solares, el radio de Schwarzschild es 10^9 veces mayor, de modo que en $r = 2M$ la curvatura y las fuerzas de marea todavía no son muy grandes. En este caso, el radio crítico de supervivencia está a unos 200.000 km de la singularidad, es decir a unos $66 \cdot 10^{-6}$ veces el radio de Schwarzschild. En general, el tiempo propio máximo que un observador tarda en llegar a la singularidad desde el radio de Schwarzschild de un agujero negro con masa m es

$$\tau_{\max} = 1,5 \cdot 10^{-5} \frac{m}{m_\odot} \; s, \tag{12.36}$$

donde $m_\odot$ es la masa del Sol. O sea, unos $10^{-5}s$ para un agujero negro estelar, pero unas cómodas 4 horas y pico para un agujero negro supermasivo de 10^9 masas solares.

Capítulo 13

Cosmología relativista

Cuando iba a dar un seminario en Japón, me pidieron que no mencionara el posible recolapso de universo, porque podía afectar a la bolsa.
(S. Hawking, 1996)

13.1. Relatividad general y cosmología

La cosmología, el estudio del universo entero, es casi tan antigua como la civilización humana. Ya los filósofos presocráticos, como Thales de Mileto o Anaximandro (siglo VI A.C.) pensaban poder deducir ciertas afirmaciones sobre la forma y la constitución del universo. Sin embargo, la cosmología sólo ha llegado a formar parte de la física cuando, gracias a la relatividad general, se concibió el universo como un sistema dinámico, regido por las mismas leyes físicas que rigen la materia dentro del universo. La relatividad general predice una estrecha relación entre la estructura del espaciotiempo y su contenido de materia y energía. Es por lo tanto un paso lógico intentar utilizar las ecuaciones de Einstein para estudiar la dinámica del universo entero: su forma, su contenido y su evolución.

Obviamente es imposible encontrar una solución exacta de las ecuaciones de Einstein que describe el universo entero con todas las estrellas, galaxias, ondas gravitacionales, campos eléctricos y magnéticos en todo detalle. Pero esto tampoco es lo que pretende la cosmología: la cosmología describe la dinámica del universo entero, a muy grandes escalas, donde la influencia de galaxias individuales e incluso cúmulos de galaxias son meramente perturbaciones despreciables.

Aún así parece que para poder afirmar algo más o menos razonable sobre la forma o la evolución del universo, hace falta información bastante detallada sobre la distribución de la materia o las condiciones iniciales, cosas difícilmente obtenibles. Sin embargo, la cosmología relativista es capaz de sacar modelos muy realistas gracias a dos principios básicos.

El primero es el *Principio Cosmológico*, que pretende decir algo sobre la forma del universo. La idea es que, sea como sea la evolución del universo, en cualquier momento su aspecto es el mismo en todos los puntos y en todas las direcciones. Este aspecto puede variar a lo largo de la evolución, pero por lo menos varía del mismo modo en todos los sitios. Más preciso, el Principio Cosmológico se formula como:

Principio Cosmológico: *En cualquier momento, el universo es homogéneo e isótropo a muy grandes escalas.*

Por lo tanto, el Principio Cosmológico implica que la métrica del universo se puede escribir como una familia de hipersuperficies (superficies tridimensionales) espaciales,[1] cada una homogénea e isótropa, que representan el universo a un tiempo t constante y juntos describen la evolución en el tiempo.

Homogéneo significa que todos los puntos de la superficie son equivalentes, no hay ningún punto privilegiado. Matemáticamente esto implica que la métrica de la superficie tiene tanta simetría que es posible relacionar cualesquiera dos puntos de la superficie a través de una transformación de simetría. Por otro lado, la isotropía de las superficies quiere decir que no hay ninguna dirección priviligiada, que las superficies tienen el mismo aspecto en todas las direcciones. Matemáticamente esto implica que la métrica de la superfice es esféricamente simétrica.

Nótese que la homogeneidad es una propiedad global de una variedad, mientras que la isotropía es una propiedad relacionada con un punto específico. En principio la homogeneidad y la isotropía son dos propiedades independiendes: un cilindro es homogéneo, pero no isótropo, mientras que un cono es isótropo visto desde el vértice, pero no homogéneo. Sin embargo, si un espacio es isótropo desde cualquier punto (es decir globalmente isótropo), entonces también es homogéneo. Y vice versa, si es isótropo desde un punto particular y homogéneo, entonces es globalmente isótropo. Un espacio que es homogéneo e isótropo es *máximamente simétrico*, es decir, tiene el número máximo de simetrías. Matemáticamente, las variedades que son máximamente simétricas son *espacios con curvatura constante*, una propiedad que se refleja en la siguiente condición sobre el tensor de Riemann,

$$R_{\mu\nu\rho\lambda} = K\Big(g_{\mu\lambda}g_{\nu\rho} - g_{\mu\rho}g_{\nu\lambda}\Big), \tag{13.1}$$

donde K es una constante con dimensión $\mathbf{L}^{-2}$, que está relacionada con el radio de curvatura.[2]

El Principio Cosmológico resume por lo tanto lo que ya habíamos dicho antes: que a escalas cosmológicas el universo tiene el mismo aspecto en todos los sitios, ya que las fluctuaciones y perturbaciones locales están promediadas a estas escalas. En cierto modo el Principio Cosmológico es una generalización del Principio Copernicano: mientras que Nicolaus Copernico (1473 - 1543) afirmaba en su *Revolutionibus* que la Tierra no ocupa ningún lugar preferido en el sistema solar, el Principio Cosmológico lo afirma para cualquier punto del Universo. Originalmente, más que un principio físico, el Principio Cosmológico era una conjetura de simplicidad. Como veremos en la siguiente sección, este principio fija casi completamente la forma de la métrica. Si el Principio Cosmológico no fuera verdad, la cosmología relativista sería mucho más difícil de tratar.

Aún así cabe preguntarse hasta que punto el Principio Cosmológico es cierto. Sabemos por las observaciones que las estrellas están concentradas en galaxias, éstas en cúmulos de galaxias que a su vez forman supercúmulos con grandes vacíos entremedios, a escalas de 10^6 años luz, así que a primera vista no parecen satisfacer las condiciones exigidas por el Principio Cosmológico. Sin embargo observaciones con radioondas y rayos X cósmicos indican que el universo efectivamente es bastante homogéneo a escalas de 10^9 años luz. Pero la mejor indicación del la veracidad del Principio Cosmológico llegó en 1965, cuando Arno Penzias (1933 -) y Robert W. Wilson (1936 -) descubrieron la radiación cósmica de fondo de microondas, correspondiendo a la radiación térmica, proveniente de un cuerpo negro con una temperatura de $T = 2,7K$. Esta radiación cósmica de fondo es en realidad el residuo de la radiación térmica de un pasado mucho más caliente del universo y fue predicha por el físico ruso George Gamov (1904 - 1968) en 1948, como

[1]Una descripción de un espacio en términos de hipersuperficies (no necesariamente espaciales), tal que cada punto del espacio esta situado en exactamente una hipersuperficie, se llama una *foliación*. Los Principios Cosmológicos dicen que los espacios que describen solucones cosmológicas son foliaciones con secciones espaciales de curvatura constante.

[2]El análisis dimensional nos dice que K es básicamente R_0^{-2}. Sin embargo, el signo de K varía de caso en caso. Nuestros convenios del tensor de Riemann, establecidos en el Capítulo 7, son tales que para el caso riemanniano, la esfera N-dimensional $\mathbb{S}^N$, el espacio con curvatura constante positiva, tiene $K > 0$ y el hiperboloide $\mathbb{H}^N$, con curvatura constante negativa tiene $K < 0$. Sin embargo para el caso lorentziano es justo al revés: el espacio de De Sitter (13.67), con curvatura constante positiva, tiene $K < 0$ y el espacio de anti-De Sitter (13.70), con curvatura constante negativa, tiene $K > 0$. Disculpamos por las molestias.

una consecuencia directa del modelo del Big Bang. La radiación cósmica de fondo, que nos proporciona información sobre cuando el universo todavía era muy joven, resulta ser muy isótropa, confirmando de manera extraordinaria el Principio Cosmológico. En realidad no fue hasta 1992 cuando el satelite COBE logró medir las primeras anisotropías en la radiación de fondo, cuyas fluctuaciones son sólo del orden de $\Delta T/T \sim 10^{-5}$. En el Capítulo **??** haremos una discusión más extendida de la radición de fondo cósmica y lo que nos enseña sobre el universo temprano.

El segundo principio básico de la cosmología relativista es el *Postulado de Weyl*, que intenta modelar el contenido de materia del universo. Igual que el Principio Cosmológico afirma que las fluctuaciones de densidad son muy pequeñas a escala cosmológica, el Postulado de Weyl dice que las velocidades propias de la materia son pequeñas en comparación con el movimiento cosmológico.

Postulado de Weyl: *La materia a escalas cosmológicas se comporta como un fluido perfecto, cuyas componentes se mueven a lo largo de geodésicas temporales, que no se intersectan, salvo (posiblemente) en un punto en el pasado.*

También el Postulado de Weyl se ve satisfecho en las observaciones: aunque sin duda existen interacciones gravitatorias entre las distintas galaxias, que en ocasiones llevan las galaxias a colisionarse y mezclarse, en general las velocidades particulares causadas por estas interacciones son despreciables con respecto a las velocidades generadas por la evolución del universo.

El Postulado de Weyl supone una clase de observadores priviligiados en el universo: los que están en resposo con respecto al fluido perfecto y cuyo movimiento por lo tanto únicamente está determinado por la evolución del universo. A estos observadores se les suele llamar *observadores comóviles*. También podemos definir un *tiempo cosmológico*, siendo la direccion temporal de un observador comóvil. Este tiempo cosmológico será útil para describir la evolución del universo y calcular su edad.

13.2. La métrica de Friedmann-Robertson-Walker

El Principio Cosmológico y el Postulado de Weyl determinan casi por completo la forma de la métrica del espaciotiempo. Por un lado, el Postulado de Weyl implica que se puede foliar el espaciotiempo con una familia de hipersuperficies espaciales, que son las superficies de simultaneidad $t = \mathrm{cte}$ con respecto al tiempo cosmológico t. Y por otro lado, el Principio Cosmológico dicta que estas superficies han de ser máximalmente simétricas. El Anstaz para la métrica por lo tanto se puede escribir sin pérdida de generalidad como

$$ds^2 \;=\; dt^2 \;-\; S^2(t)\, \tilde{g}_{ij}(x) dx^i dx^j, \tag{13.2}$$

donde $\tilde{g}_{ij}$ es la métrica de las secciones espaciales tridimensionales con curvatura constante. En otras palabras, el tensor de Riemann $\tilde{R}_{ijkl}$ de la métrica $\tilde{g}_{ij}$ satisface la condición

$$\tilde{R}_{ijkl} = K\Big(\tilde{g}_{il}\tilde{g}_{jk} - \tilde{g}_{ik}\tilde{g}_{jl}\Big). \tag{13.3}$$

La función $S^2(t)$ es el *factor de escala*, una función del tiempo cosmológico que mide la expansión o la contracción del universo (o, para ser más exactos, de las secciones espaciales). Obsérvese que para que las secciones espaciales sean homogéneas e isótropas en todo momento, todas las direcciones espaciales tienen que evolucionar de la misma manera y por lo tanto el factor de escala tiene que ir multiplicando a todas las direcciones espaciales. En principio es posible incluir una función del tiempo $f^2(t)$ multiplando el término dt^2, pero es fácil ver que se podría absorber esta función con una redefinición de la coordenada temporal. La métrica (13.2) con la condición (13.3) es por lo tanto la métrica más general de un universo homogéneo e isótropo.

El problema central de la cosmología relativista es determinar las funciones $S^2(t)$ y $\tilde{g}_{ij}(x)$ en el Ansatz (13.2) en función del contenido de energía y materia del universo. El factor de escala $S^2(t)$ se determinará a través de las ecuaciones de Einstein, ya que éstas describen la dinámica del sistema. Nos ocuparemos de este problema en las siguientes secciones. Sin embargo, hallar $\tilde{g}_{ij}(x)$ es un problema puramente geométrico, puesto que implica resolver la ecuación (13.3). Dedicaremos el resto de esta sección a encontrar e interpretar las soluciones de esta ecuación.

La isotropía del espacio implica una simetría esférica, por lo tanto podemos escribir la métrica en las secciones espaciales como (véase sección 12.1)

$$d\tilde{s}^2 = e^{2B(\bar{r})} d\bar{r}^2 + \bar{r}^2 \Big(d\theta^2 + \sin^2\theta d\varphi^2 \Big), \tag{13.4}$$

donde $B(\bar{r})$ es una función aún desconocida de la coordenada radial $\bar{r}$. En lugar de sustituir este Ansatz en la ecuación (13.3), vamos a determinar la función $B(\bar{r})$ a través de la ecuación para el tensor de Ricci:

$$\tilde{R}_{ik} = -2K\tilde{g}_{ik}. \tag{13.5}$$

En un espacio tridimensional (como es nuestro caso) las condiciones (13.3) y (13.5) son equivalentes,[3] por lo tanto es preferible resolver la última, ya que es más sencillo calcular el tensor de Ricci que el de Riemann. Sin embargo en general la condición (13.5) es claramente más debil que la condición (13.3). Todas las métricas de curvatura constante satisfacen la ecuación (13.5), pero no todas las métricas que satisfacen (13.5) tienen curvatura constante. Las métricas que satisfacen la ecuación (13.5) se llaman *métricas tipo Einstein*.[4]

Los símbolos de Christoffel no nulos del Ansatz (13.4) son (ejerc. ó compárese con (12.4))

$$\begin{aligned} &\tilde{\Gamma}_{\bar{r}\bar{r}}^{\bar{r}} = B', && \tilde{\Gamma}_{\bar{r}\theta}^{\theta} = \tilde{\Gamma}_{\bar{r}\varphi}^{\varphi} = \tfrac{1}{\bar{r}}, \\ &\tilde{\Gamma}_{\theta\theta}^{\bar{r}} = -\bar{r}\, e^{-2B}, && \tilde{\Gamma}_{\varphi\varphi}^{\theta} = -\sin\theta\cos\theta, \\ &\tilde{\Gamma}_{\varphi\varphi}^{\bar{r}} = -\bar{r}\sin^2\theta\, e^{-2B}, && \tilde{\Gamma}_{\theta\varphi}^{\varphi} = \operatorname{cotg}\theta, \end{aligned} \tag{13.6}$$

de modo que las componentes no-triviales del tensor de Ricci vienen dadas por (ejerc.)

$$\tilde{R}_{\bar{r}\bar{r}} = -\frac{2B'}{\bar{r}}, \qquad \tilde{R}_{\theta\theta} = -1 + e^{-2B} - \bar{r}B'e^{-2B}, \qquad \tilde{R}_{\varphi\varphi} = \sin^2\theta\, \tilde{R}_{\theta\theta}, \tag{13.7}$$

donde la prima denota la derivada con respecto a $\bar{r}$. La ecuación (13.5) se reduce entonces, en nuestro caso, a dos ecuaciones independientes

$$\frac{B'}{\bar{r}} = Ke^{2B}, \qquad -e^{-2B}\Big(1 - \bar{r}B'\Big) + 1 = 2K\bar{r}^2, \tag{13.8}$$

que tienen como solución

$$e^{2B} = \frac{1}{1 - K\bar{r}^2}. \tag{13.9}$$

La métrica de una superficie (tridimensional) con curvatura constante viene dada por lo tanto por

$$d\tilde{s}^2 = \frac{1}{1 - K\bar{r}^2} d\bar{r}^2 + \bar{r}^2 \Big(d\theta^2 + \sin^2\theta d\varphi^2 \Big), \tag{13.10}$$

[3]No es directo ver que en tres dimensiones las dos condiciones son equivalentes. La manera más fácil es observar que al resolver (13.5) todas las constantes de integración (y por lo tanto la solución entera) quedan determinadas. Una manera más elegante es darse cuenta de que en tres dimensiones tanto el tensor de Riemann como el de Ricci tienen 6 grados de libertad y por lo tanto los dos llevan exactamente la misma información sobre la curvatura del espacio. En otras palabras, la ecuación (13.3) no impone ninguna restricción más que la ecuación (13.5). En dimensiones mayores que 3 (es decir, en universos 5-dimensionales o más), sí es necesario resolver la ecuación (13.3).

[4]Obsérvese que la relación entre los espacios de curvatura constante y los espacios tipo Einstein es un generalización de la relación entre el espacio plano y un espacio Ricci-plano. En realidad el espacio plano y Ricci-plano no son nada más que un espacio de curvatura constante y un espacio de Einstein con $K = 0$.

donde la constante K puede tener un valor arbitrario positivo, negativo o cero, correspondiendo respectivamente a una variedad tridimensional con curvatura constante positiva, negativa o cero. Para interpretar esta métrica y para futuro comodidad es conveniente sacar un factor común $|K|^{-1}$, a través del rescaleo de la coordenada radial $\bar{r} = r/\sqrt{|K|}$ (para $K \neq 0$). La métrica (13.10) entonces coge la forma

$$d\tilde{s}^2 = |K|^{-1}\Big[\frac{1}{1-kr^2}dr^2 + r^2\Big(d\theta^2 + \sin^2\theta d\varphi^2\Big)\Big], \qquad \text{para } K \neq 0$$

$$d\tilde{s}^2 = dr^2 + r^2\Big(d\theta^2 + \sin^2\theta d\varphi^2\Big), \qquad \text{para } K = 0 \qquad (13.11)$$

donde ahora la constante k está definida como $k = K/|K|$.

Para interpretar la métrica (13.11), hay que considerar uno por uno los tres casos de K positivo, negativo o cero. El caso más sencillo es sin duda $K = 0$: en (13.11) reconocemos directamente la métrica para $\mathbb{R}^3$ en coordenadas esféricas. Esto era de esperar, ya que para $K = 0$ la ecuación (13.3) se reduce a la condición para el espacio plano. Intuitivamente sabemos que $\mathbb{R}^3$ es un espacio de curvatura constante, puesto que tiene curvatura cero en todos los puntos.

El caso $K > 0$, es decir curvatura constante positiva, es un poco más sutil. Nótese que ahora $k = 1$ y por lo tanto el rango de la coordenada r cubre sólo el intervalo $]-1, 1[$, ya que la componente $\tilde{g}_{rr}$ se vuelve singular cuando $r \to \pm 1$. Es por lo tanto natural hacer el cambio de coordenadas

$$r = \sin\chi \qquad \Longleftrightarrow \qquad d\chi = \frac{dr}{\sqrt{1-r^2}}, \qquad (13.12)$$

de modo que la métrica (13.11) se convierte en

$$d\tilde{s}^2 = K^{-1}\Big[d\chi^2 + \sin^2\chi\Big(d\theta^2 + \sin^2\theta d\varphi^2\Big)\Big]. \qquad (13.13)$$

Esta métrica es la de una esfera tridimensional $\mathbb{S}^3$ con radio $K^{-\frac{1}{2}}$. La manera más fácil verlo es embeber (13.13) en el espacio plano cuadridmensional $\mathbb{R}^4$ a través de las coordenadas

$$\begin{aligned} X^1 &= K^{-\frac{1}{2}}\ \sin\chi\ \sin\theta\ \cos\varphi, \qquad & X^3 &= K^{-\frac{1}{2}}\ \sin\chi\ \cos\theta, \\ X^2 &= K^{-\frac{1}{2}}\ \sin\chi\ \sin\theta\ \sin\varphi, \qquad & X^4 &= K^{-\frac{1}{2}}\ \cos\chi. \end{aligned} \qquad (13.14)$$

Claramente estas coordenadas satisfacen la ligadura $(X^1)^2 + (X^2)^2 + (X^3)^2 + (X^4)^2 = K^{-1}$, de modo que las coordenadas $\{\chi, \theta, \varphi\}$ efectivamente describen una tres-esfera en $\mathbb{R}^4$ con radio $K^{-\frac{1}{2}}$. Además la métrica (13.13) corresponde a la métrica de esta tres-esfera, porque sustituyendo la parametrización (13.14) en la métrica cartesiana de $\mathbb{R}^4$ (¡sin olvidarse de la ligadura!) obtenemos (ejerc.)

$$ds^2 = (dX^1)^2 + (dX^2)^2 + (dX^3)^2 + (dX^4)^2 = K^{-1}\Big[d\chi^2 + \sin^2\chi\Big(d\theta^2 + \sin^2\theta d\varphi^2\Big)\Big]. \qquad (13.15)$$

El último caso es la variedad con curvatura constante negativa, $K < 0$, o equivalentemente $k = -1$. En este caso podemos hacer el cambio de coordenadas

$$r = \sinh\chi \qquad \Longleftrightarrow \qquad d\chi = \frac{dr}{\sqrt{1+r^2}}, \qquad (13.16)$$

de modo que la métrica (13.11) se convierte en

$$d\tilde{s}^2 = K^{-1}\Big[d\chi^2 + \sinh^2\chi\Big(d\theta^2 + \sin^2\theta d\varphi^2\Big)\Big]. \qquad (13.17)$$

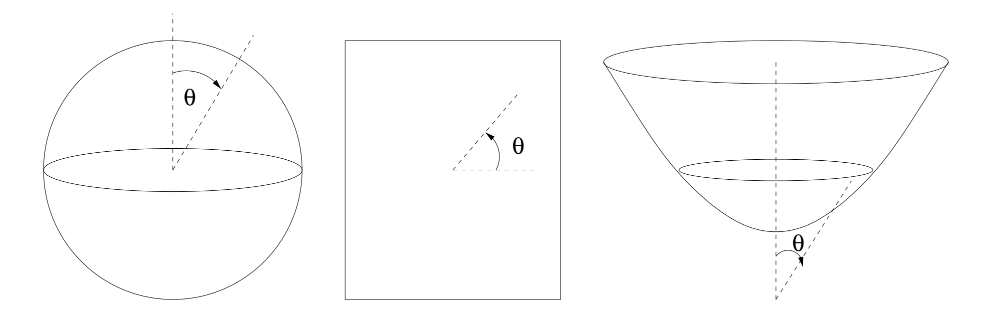


Figura 13.1: *Los tres espacios con curvatura constante (aquí en su versión bidimensional): la esfera $\mathbb{S}^N$ con curvatura positiva (izquierda), el plano $\mathbb{R}^N$ con curvatura cero (centro) y el hiperboloide $\mathbb{H}^N$ con curvatura negativa (derecha), cada uno con su angulo azimutal θ. La esfera $\mathbb{S}^N$ se puede embeber en el espacio euclideo $\mathbb{R}^{N+1}$, mientras que el hiperboloide $\mathbb{H}^N$ se puede embeber en el espacio lorentiano $\mathbb{R}^{1,N}$.*

Esta métrica describe un hiperboloide tridimensional $\mathbb{H}^3$, como que se puede comprobar considerando la parametrización

$$X^1 = K^{-\frac{1}{2}} \sinh\chi \, \sin\theta \, \cos\varphi, \qquad X^3 = K^{-\frac{1}{2}} \sinh\chi \, \cos\theta,$$
$$X^2 = K^{-\frac{1}{2}} \sinh\chi \, \sin\theta \, \sin\varphi, \qquad X^0 = K^{-\frac{1}{2}} \cosh\chi. \tag{13.18}$$

El hiperboloide $\mathbb{H}^3$ no se puede embeber en $\mathbb{R}^4$, sino en su versión lorentziana, el espacio de Minkowski $\mathbb{R}^{1,3}$: efectivamente, las coordenadas satisfacen la ligadura

$$(X^0)^2 - (X^1)^2 - (X^2)^2 - (X^3)^2 = K^{-1}, \tag{13.19}$$

lo que determina en $\mathbb{R}^{1,3}$ una superficie a distancia temporal constante $1/\sqrt{|K|}$ del origen (véase la figura 13.1). La métrica (13.17) se obtiene sustituyendo la parametrización (13.18) y la ligadura en la métrica $ds^2 = (dX^0)^2 - (dX^1)^2 - (dX^2)^2 - (dX^3)^2$ (aunque con signo opuesto por ser una superficie temporal en un espacio lorentziano).

En resumen, las tres superficies tridimensionales con curvatura constante son por lo tanto la esfera $\mathbb{S}^3$ (curvatura positiva), el plano $\mathbb{R}^3$ (curvatura cero) y el hiperboloide $\mathbb{H}^3$ (curvatura negativa). No es sorprendente que fue en estos espacios donde históricamente se desarolló la geometría diferencial, dado que estos son los casos con más simetría: la geometría plana en $\mathbb{R}^3$ de Euclides, la geometría esférica por la cartografía y astronomía y la geometría no-euclidea de Bolyai y Lobachevsky para el hiperboloide en el siglo XIX.

Por lo tanto, sustituyendo la forma de la métrica (13.11) en nuestro Ansatz cosmológico (13.2), vemos que la métrica de un universo homogéneo e isótropo siempre se puede escribir como

$$ds^2 = dt^2 - a^2(t)\Big[\frac{1}{1-kr^2}dr^2 + r^2\Big(d\theta^2 + \sin^2\theta d\varphi^2\Big)\Big], \tag{13.20}$$

donde hemos absorbido el radio $|K|$ de los espacios tridimensionales en un nuevo factor de escala $a(t)$, definido como

$$a(t) = |K|^{-\frac{1}{2}} \, S(t) \qquad \text{para } K \neq 0,$$
$$a(t) = S(t) \qquad \text{para } K = 0. \tag{13.21}$$

La imagen por lo tanto es que a cualquier momento $t = t_0$, las secciones espaciales son superficies de curvatura constante y el factor de escala $a(t)$ representa de cierto modo el "tamaño" de esta superficie espacial. En el caso de $k = 1$, la función $a(t)$ es el radio de la tres-esfera en el momento t y, aunque para los otros dos casos es más difícil visualisar, el aumento o disminución del factor

de escala implica una expansión o contracción de la sección espacial y marca por lo tanto una escala en las secciones espaciales. La expansión del universo por lo tanto no se corresponde con la imagen de diferentes galaxias alejándose unas de otras debido a los movimientos radiales de cada una de ellas en un espacio fijo, sino más bien a la imagen del contenido de materia y energía diluyéndose, debido a la constante creación de espacio, con el aumento del factor de escala. Obsérvese que para $k = 0$, el factor de escala es adimensional, pero para $k = \pm 1$, $a(t)$ tiene dimensiónes **L**.

La métrica (13.20) se llama la *métrica de Friedmann-Robertson-Walker* (FRW), llamada así por el físico-meteorólogo ruso Alexander Friedmann (1888 - 1925), el físico americano Howard Robertson (1903 - 1961) y el matemático inglés Arthur Walker (1909 - 2001). Friedmann propuso en 1922 la métrica (13.20) como Ansatz para el universo, dedujo las ecuaciones de Friedmann y obtuvo una de las primeras soluciones realistas de un universo en expansión.[5] En 1935 y 1936 Robertson y Walker demostraron independientemente que la métrica que propuso Friedmann es el Ansatz más general que describe un universo homogéneo e isótropo. Las coordenadas en que está escrita la métrica de FRW en (13.20) se suelen llamar *coordenadas comóviles*, ya que el tiempo t es el tiempo propio de un observador que se mueve con la expansión del universo.

13.3. Las ecuaciones de Friedmann

Ya que tenemos la forma general de la métrica de un universo homogéneo e isótropo, podemos concentrarnos en el verdadero problema de la cosmología relativista: la evolución del universo, codificado en la dinámica del factor de escala $a(t)$. Para esto hay que resolver la ecuación de Einstein, utilizando el Ansatz que acabamos de derivar y la descripción apropiada de la materia.

Para calcular los tensores de curvatura de la métrica (13.20), conviene escribirla como

$$ds^2 = dt^2 - a^2(t)\, \tilde{g}_{ij}(x)\, dx^i dx^j, \tag{13.22}$$

donde $\tilde{g}_{ij}(x)$ es la métrica de las secciones espaciales de curvatura constante, que hemos calculado antes. No sólo de esta forma podemos tratar los tres casos $k = -1, 0, 1$ simultáneamente, sino también resulta que el resultado es explicitamente independiente de las coordenadas utilizadas en las secciones espaciales, ya que lo único que necesitamos saber es que

$$\tilde{R}_{ijkl} = k(\tilde{g}_{il}\tilde{g}_{jk} - \tilde{g}_{ik}\tilde{g}_{jl}), \qquad \tilde{R}_{ij} = -2k\tilde{g}_{ij}, \qquad \tilde{R} = -6k. \tag{13.23}$$

Un cálculo rutinario revela que los símbolos de Christoffel non-cero son

$$\Gamma^t_{ij} = a\dot{a}\,\tilde{g}_{ij}, \qquad \Gamma^i_{tj} = \frac{\dot{a}}{a}\,\delta^i_j, \qquad \Gamma^k_{ij} = \tilde{\Gamma}^k_{ij}, \tag{13.24}$$

donde el punto indica derivar con respecto a la coordenada t y los $\tilde{\Gamma}^k_{ij}$ son los símbolos de Christoffel (13.6) de la métrica $\tilde{g}_{ij}$. Del mismo modo el tensor de Ricci, el escalar de Ricci y el tensor de Einstein vienen dados por

$$R_{tt} = 3\frac{\ddot{a}}{a}, \qquad R_{ij} = -\Big[2k + a\ddot{a} + 2\dot{a}^2\Big]\tilde{g}_{ij}, \qquad R = 6\Big[a^{-2}k + \frac{\ddot{a}}{a} + \Big(\frac{\dot{a}}{a}\Big)^2\Big],$$
$$G_{tt} = -3\Big[a^{-2}k + \Big(\frac{\dot{a}}{a}\Big)^2\Big], \qquad G_{ij} = \Big[k + 2a\ddot{a} + \dot{a}^2\Big]\tilde{g}_{ij}. \tag{13.25}$$

[5] En 1927, dos años después de la muerte de Friedmann, el sacerdote belga George Lemaître (1894 - 1966) derivó independientemente las ecuaciones de Friedmann y presentó varias soluciones cosmológicas. Por eso a veces se le denomina (13.20) como la *métrica de Friedmann-Lemaître-Robertson-Walker* (FLRW).

Aunque no es estrictamente necesario para resolver las ecuaciones de Einstein, conviene calcular también el tensor de Riemann, para estudiar las singularidades de las soluciones que vamos a encontrar. Las únicas componentes que no son cero son de la forma

$$R_{titj} \;=\; -a\ddot{a}\,\tilde{g}_{ij}, \qquad\qquad R_{ijkl} \;=\; -a^2\Big[k+\dot{a}^2\Big](\tilde{g}_{il}\tilde{g}_{jk}-\tilde{g}_{ik}\tilde{g}_{jl}). \tag{13.26}$$

Para determinar $a(t)$, hay que resolver un sistema de ecuaciones diferenciales para el factor de escala $a(t)$, que se obtiene sustituyendo el Ansatz de FRW (13.22) en las ecuaciones de Einstein

$$R_{\mu\nu} \;-\; \frac{1}{2}g_{\mu\nu}R \;=\; -\kappa T_{\mu\nu}, \tag{13.27}$$

donde el tensor de energía-momento contiene las contribuciones de todos los tipos de energía-momento presente en el universo (materia, radiación, constante cosmológica, ...). En la sección 13.5 discutiremos en más detalle las propiedades de los distintos tipos de materia y energía, pero de momento nos basta con saber que por el Principio de Weyl se puede considerar el contenido del universo como un fluido perfecto. Podemos escribir por lo tanto

$$T_{tt} \;=\; \rho, \qquad\qquad T_{ij} \;=\; a^2\tilde{g}_{ij}P, \tag{13.28}$$

donde $\rho = \sum_\alpha \rho_\alpha$ y $P = \sum_\alpha P_\alpha$ son respectivamente la densidad y la presión total de todos los tipos de energía y materia presentes en el universo.

En 1922 Friedmann sustituyó el Ansatz (13.22) en las ecuaciones de Einstein, obteniendo así las *ecuaciones de Friedmann* (y corrigiendo un error que comitió Einstein al derivar estas ecuaciones para su modelo del universo estático de 1917):

$$\begin{aligned} \Big(\frac{\dot{a}}{a}\Big)^2 \;&=\; \frac{1}{3}\kappa\rho \;-\; \frac{k}{a^2}, \\ \frac{\ddot{a}}{a} \;+\; \frac{1}{2}\Big(\frac{\dot{a}}{a}\Big)^2 \;&=\; -\frac{1}{2}\kappa P \;-\; \frac{k}{2a^2}. \end{aligned} \tag{13.29}$$

Aunque al conjunto de estas ecuaciones se les llama las ecuaciones de Friedmann, también a la primera ecuación sólo, la componente $\{tt\}$ de la ecuación de Einstein, se le denomina confusamente la *ecuación de Friedmann*, mientras que la segunda, la componente $\{ij\}$, se llama la *ecuación de evolución*. Se puede simplificar bastante esta última, combinándola con la ecuación de Friedmann, dando lugar a la *ecuación de aceleración*

$$\frac{\ddot{a}}{a} \;=\; -\frac{1}{6}\kappa\Big(\rho+3P\Big). \tag{13.30}$$

La ecuación de Friedmann (13.29a) relaciona por lo tanto la velocidad de expansión del universo con la densidad de energía y con la curvatura de las secciones espaciales. Obsérvese que sólo es una ecuación diferencial de primer orden y por lo tanto no es realmente una ecuación de movimiento, sino más bien una ligadura para a, $\dot{a}$, ρ y k. Por otro lado, la ecuación de evolución (13.29b) sí es una ecuación de segundo orden y actúa como la ecuación de movimiento de $a(t)$ (de allí su nombre). En la práctica muchas veces basta con resolver la ecuación de Friedmann para determinar el factor de escala. Veremos en la sección 13.5 que la ley de conservación de energía y la ecuación de Friedmann implican la ecuación de aceleración.

Finalmente, a veces conviene reescribir la métrica de FRW (13.22) en las llamadas *coordenadas conformes*, donde el factor de escala aparece en la métrica como un factor conforme. Se define el *tiempo conforme* τ, a través de una reparametrización de la coordenada temporal, como

$$d\tau \;=\; a^{-1}(t)dt \qquad \Leftrightarrow \qquad dt \;=\; a(\tau)d\tau, \tag{13.31}$$

de modo que la métrica (13.22) coge la forma

$$ds^2 = a^2(\tau)\Big[d\tau^2 - \tilde{g}_{ij}dx^i dx^j\Big], \tag{13.32}$$

donde el factor de escala $a(\tau) = a(t(\tau))$ ahora es una función del tiempo conforme. El tiempo conforme no es el tiempo propio de ningún observador en particular (obsérvese que τ tiene dimensión $\mathbf{L}^{-1}$para $k = 0$, pero es adimensional para $k = \pm 1$), pero estas coordenadas tienen algunas ventajas, como por ejemplo dejar claro que las métricas FRW con $k = 0$ son conformemente planas.

Los tensores de curvatura en estas coordenadas vienen dados por

$$\begin{aligned} R_{\tau\tau} &= 3\Big[a^{-1}a'' - a^{-2}(a')^2\Big], & G_{\tau\tau} &= -3\Big[k + a^{-2}(a')^2\Big], \\ R_{ij} &= -\Big[2k + a^{-1}a'' + a^{-2}(a')^2\Big], & G_{ij} &= \Big[k + 2a^{-1}a'' - a^{-2}(a')^2\Big], \\ R &= 3\Big[2a^{-2}k + 2a^{-3}a''\Big], \end{aligned} \tag{13.33}$$

donde el acento indica una derivación con respecto al tiempo conforme τ. Las ecuaciones de Friedmann en estas coordenadas son entonces

$$\begin{aligned} \Big(\frac{a'}{a}\Big)^2 &= \frac{1}{3}\kappa a^2\rho - k, \\ \frac{a''}{a} - \frac{1}{2}\Big(\frac{a'}{a}\Big)^2 &= -\frac{1}{2}\kappa a^2 P - \frac{1}{2}k, \end{aligned} \tag{13.34}$$

y la ecuación de acelerción viene dada por

$$\frac{a''}{a} = \frac{1}{6}\kappa a^2(\rho + 3P). \tag{13.35}$$

13.4. Distancias y horizontes cosmológicos

Hay varias maneras inequivalente de definir distancias entre dos puntos en una métrica de Friedmann-Robertson-Walker, cada uno con sus ventajas y su significado físico.

Quizá el más intuitivo es la *distancia instantánea*, o la *distancia geométrica*. Esta definición usa el hecho de que la métrica de Friedmann-Robertson-Walker tiene un tiempo cosmológico preferido, el tiempo propio de un observador comóvil, de modo que podemos calcular la distancia entre dos puntos a un tiempo t constante. Intuitivamente equivale a la longitud de una goma extendida entre dos puntos. Sin pérdida de generalidad podemos tomar uno de los dos puntos en el origen del sistema de coordenadas, de modo que el desplazamiento sea puramente radial. En otras palabras podemos tomar $dt = d\Omega_2 = 0$ en la métrica (13.20). La distancia geométrica está entonces definida como

$$D(t) = \int ds = \int_0^r \frac{a(t)\,dr'}{\sqrt{1-kr'^2}} = a(t)\int_0^\chi d\chi' = a(t)\,\chi, \tag{13.36}$$

donde en la tercera igualdad hemos usado el cambio de coordenadas (13.12) para $k = 1$, (13.16) para $k = -1$, y simplemente tomamos $\chi = r$ para $k = 0$. Vemos por lo tanto que la distancia geométrica es la distancia euclidea χ entre los dos puntos, medida dentro de las secciones espaciales, multiplicada por el fector de escala. Al expandir o contraerse el universo, las distancias geométricas entre los distintos eventos varían.

Podemos calcular la velocidad de receso de las galaxias, derivando la definición (13.36) con respecto a t:

$$v_{\rm rec} \;=\; \dot{D}(t) \;=\; \dot{a}(t)\,\chi \;=\; \frac{\dot{a}(t)}{a(t)}\,D(t). \tag{13.37}$$

Evaluada en la actualidad $t = t_0$, esta ecuación se convierte en la *Ley de Hubble*,

$$v_{\rm rec} \;=\; H_0\,D(t), \tag{13.38}$$

publicada por el astrónomo americano Edwin P. Hubble (1889 - 1953) en 1929. La ley de Hubble establece una relación lineal entre la distancia geométrica $D(t)$ a la que se encuentran las galaxias y su velocidad de receso $v_{\rm rec}$, donde la constante de proporcionalidad es la *constante de Hubble* $H_0 = \dot{a}(t_0)/a(t_0)$. Las observaciones de Hubble, que dieron lugar a su ley empírica, jugaron un importantísimo papel en la aceptacion de que nuestro universo está realmente en expansión, en contraste con lo que la humanidad siempre había creído. Obsérvese que la ley de Hubble no es una propiedad específica de nuestro universo, sino que es general para cualquier modelo cosmológico derivado de la métrica de Friedmann-Robertson-Walker.

En realidad, la relación lineal entre la distancia y la velocidad de receso solamente es válida para distancias pequeñas, a orden cero en el desarrollo de Taylor de $a(t)$. En general la relación entre ambos está determinada por el *parámetro de Hubble*

$$H(t) \;=\; \frac{\dot{a}(t)}{a(t)}, \tag{13.39}$$

que mide la velocidad de expansión (o contracción) en comparación con la escala del universo. En particular la evolución del parámetro de Hubble con el tiempo, $H(t)$, codificada en la variación del parámetro de Hubble como función de la distancia (observacionalmente medible como función del corrimiento hacia el rojo, $H(z)$), es una fuente muy valiosa de información sobre la evolución de nuestro universo.[6]

Como tiene dimensión $\mathbf{L}^{-1}$, el parámetro de Hubble marca una escala de tiempo y de distancia para el modelo cosmológico en consideración. El *radio de Hubble* $R_H(t) = H^{-1}(t)$ es la distancia a la que la velocidad de receso de las galaxias es igual a la velocidad de la luz, $v_{\rm rec} = 1$, y galaxias más allá del radio de Hubble tienen velocidades de receso superlumínicas, como se puede ver en la ley de Hubble general (13.37).[7] Por otro lado el *tiempo de Hubble* $T_H(t) = H^{-1}(t)$ es la edad que tendría el universo si se hubiera expandido siempre con la misma velocidad que ahora, y es el tiempo que necesita la luz para viajar la distancia de un radio de Hubble. Obsérvese que el tiempo de Hubble no es (necesariamente) la edad real del universo (de hecho sólo lo es para un modelo específico), ni el radio de Hubble es la frontera del universo visible, como veremos en breve.

En la práctica la distancia geométrica (13.36) no es muy útil, ya que no es una cantidad observable: no podemos ver dónde están las galaxias en este instante, sólo donde estaban cuando emitieron la luz que ahora nos llega. Podemos por lo tanto definir la *distancia de cono de luz* como la distancia espacial que ha recorrido la luz entre emisión y observación o, equivalentemente, el tiempo que ha tardado en llegar hasta nosotros. Dado que la luz sigue curvas nulas en el espaciotiempo cuadrimensional, tenemos para la métrica (13.20) con $ds = d\Omega_2 = 0$, que

$$0 \;=\; dt^2 \;-\; a^2(t)\,d\chi^2, \tag{13.40}$$

[6] El parámetro z del corrimiento hacia el rojo está definido como $1+z = \lambda_d/\lambda_e$, donde $\lambda_{e,d}$ son las longitudes de onda en el momento de emisión y detección. Para el caso cosmológico, donde la longitud de onda cambia con la expansión del universo, tenemos que $1+z = \lambda_d/\lambda_e = a(t_0)/a(t_e)$.

[7] Esto no contradice los postulados de la relatividad especial, ya que su velocidad es sublumínica para un observador inercial local (es más, será cero si la galaxia no tiene una velocidad particular). En otras palabras, la velocidad de receso (13.37) no es la velocidad con la que la galaxia se mueve por el espacio, sino la medida en que aumenta su distancia hacia nosotros debido a la expansión del espacio. Recuerda que la velocidad "de movimiento" (no de receso) sólo está definida localmente, en el espacio tangente en el punto donde se encuentra la galaxia, y no hay ningún observador que medirá velocidades superlumínicas es su entorno local.

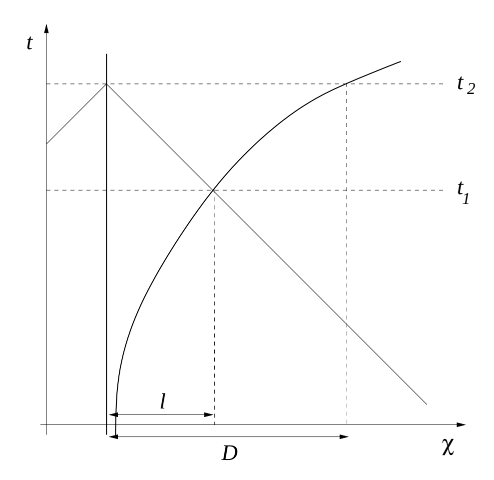


Figura 13.2: *La distancia geométrica y la distancia de cono de luz en una métrica del tipo Friedmann-Robertson-Walker: La distancia geométrica $D(t_2)$ entre dos galaxias es la distancia física que separa las galaxias a $t = t_2$ constante. La distancia de cono de luz $\ell(t_2)$ es la distancia espacial que ha viajado la luz emitida por una galaxia en $t = t_1$ hasta ser detectada por otra en $t = t_2$.*

de modo que la distancia de cono de luz, $\ell = \int_0^\chi d\chi'$, viene dada por

$$\ell \;=\; \int_{t_e}^{t_d} \frac{dt}{a(t)}. \tag{13.41}$$

Nótese que $a^{-1}(t) = d\chi/dt$ es la velocidad de la luz a distancia comóvil χ de un observador en $\chi = 0$ (la velocidad de la luz para un observador local siempre será 1), de modo que (13.41) es realmente la distancia que ha recorrido la luz entre emisión y detección.

Para cualquier $a(t)$ no-trivial, la distancia geométrica y la distancia de cono de luz dan resultados distintos. En particular la distancia geométrica siempre es finita, mientras que es posible que la integral (13.41) diverja. Si eso ocurre, quiere decir que la luz que estamos detectando viene de distancias arbitrariamente lejanas, mientras que si la integral da un resultado finito el modelo cosmológico contiene horizontes.

Ya hemos discutido brevemente los tipos de horizontes cosmológicos en la sección 10.4, pero aquí lo haremos de forma más exacta. Una solución cosmológica tiene un *horizonte de partículas* si la integral (13.41), evaluada entre el origen del universo en $t = 0$ (o en $t = -\infty$ si no tiene) y la actualidad, converge a un valor finito,

$$\ell_{\rm HP}(t) \;=\; \int_0^t \frac{dt'}{a(t')} \;<\; \infty. \tag{13.42}$$

Esto quiere decir que sólo la luz de las galaxias a distancias menores que $\ell_{\rm HP}(t)$ ha tenido tiempo para llegar hasta nosotros desde el origen del universo. Las galaxias detrás del horizonte de partículas $\ell_{\rm HP}(t)$ están demasiado lejos y su luz aún no nos ha alcanzado. El horizonte de partículas forma por lo tanto la frontera de la parte visible del universo y se está expandiendo constantemente.

De la misma manera podemos preguntarnos hasta dónde alcanza la luz que emitimos en este momento. Si la integral la integral (13.41) evaluada entre la actualidad y $t = +\infty$ (o $t = t_f$, si recolapsa en un tiempo finito),

$$\ell_{\rm HE}(t) \;=\; \int_t^\infty \frac{dt'}{a(t')} \;<\; \infty, \tag{13.43}$$

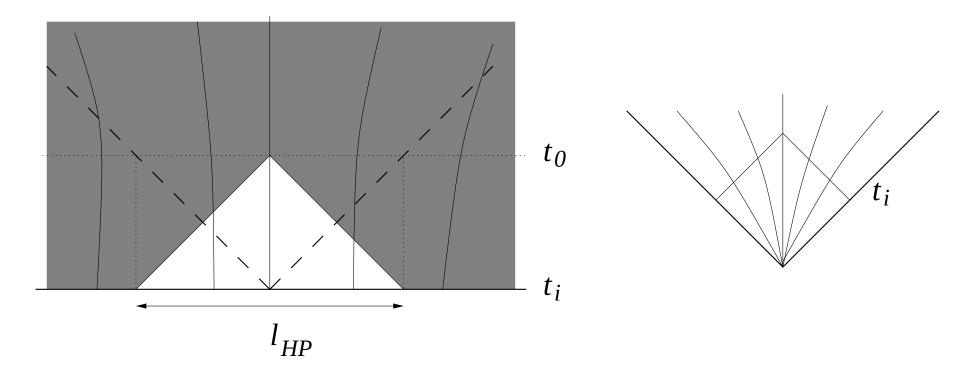


Figura 13.3: *El horizonte de partículas: Si la expansión inicial de universo es suficientemente rápida, el cono de luz pasado no intersecta las lineas de universo de todas las galaxias (izquierda) y una parte del universo es invisible para el observador (zona oscura). El tamaño del horizonte de partículas ℓ_{HP}, la frontera entre la parte visible y la parte invisible, es igual a la anchura del cono de luz pasado y crece con el tiempo (linea interrumpida). Si la expansión inicial es suficientemente lenta, el modelo no tiene horizonte de partículas, ya que el cono de luz intersecta las lineas de universo de todas las galaxias (derecha), de modo que el universo entero es observable.*

diverge, entonces las influencias causales de nuestro "aquí y ahora" puede llegar arbitrariamente lejos. Sin embargo, si la integral (13.43) converge a un valor finito, nunca (más) podremos influenciar eventos más allá del *horizonte cósmico de eventos* $\ell_{\rm HE}(t)$.

Finalmente es útil definir una medida para la aceleración o deceleración de la expansión del universo. Más concretamente se define el *parámetro de deceleración* como

$$q(t) \;=\; -\frac{\ddot{a}\, a}{\dot{a}^2}. \tag{13.44}$$

Obsérvese que q es un parámetro adimensional, que indica la deceleración del universo para valores positivos de q y aceleración para q negativo.

Podemos ilustrar estos conceptos con unos ejemplos concretos. El primer ejemplo es trivial, pero sirve para aplicar los conceptos a un caso conocido: el espacio de Minkowski, como métrica de Friedmann-Roberson-Walker con $a(t) = 1$. El parámtro de Hubble y el parámentro de deceleración son idénticamente cero y por lo tanto el radio de Hubble es infinito, ya que este universo no tiene evolución. Esto también conlleva que a la distancia geométrica y la distancia de cono de luz de un punto al origen coincide: $D(t) = \ell(t) = r$. Es fácil ver que los integrales (13.42) y (13.43) divergen, lo que confirme lo que ya sabíamos: el cono de luz en el espacio de Minkowski puede alcanzar tamaños arbitrariamente grandes, de modo que no hay ni horizonte de partículas, ni horizonte de eventos. Veremos en el Capítulo **??** que el pasado y el futuro del espacio de Minkwoski tiene una estructura causal como la parte derecha de las Figuras 13.3 y 13.4.

Un ejemplo un poco menos trivial son los modelos cosmológicos donde el factor de escala es una potencia del tiempo cosmológico, $a(t) = A_0 t^n$ para algún $n > 0$ y con A_0 una constante de proporcionalidad poco relevante ahora. Veremos en la sección 13.6 que una clase de universos espacialmente planos se comportan de esta manera. Es fácil ver que el parámetro de Hubble y el de deceleración vienen dados por

$$H \;=\; nt^{-1}, \qquad\qquad q \;=\; \frac{1-n}{n}, \tag{13.45}$$

de modo que estos modelos cosmológicos están decelerados si $0 < n < 1$ y acelerados si $n > 1$. Nótese que el radio de Hubble crece linealmente en todos los casos: $R_H = n^{-1}t$. La distancia

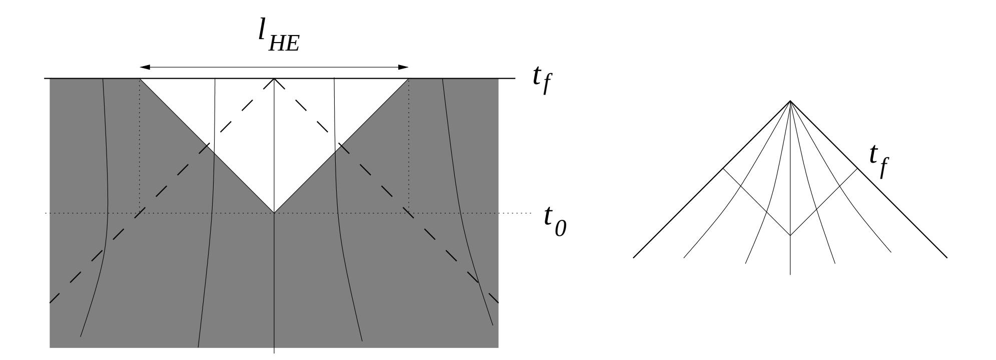


Figura 13.4: *El horizonte cósmico de eventos: Si la expansión de universo es suficientemente rápida, o si el universo recolapsa, el cono de luz futuro no intersecta las lineas de universo de todas las galaxias (izquierda) y hay partes de universo a las que las influencias causales del observador no llegan (zona oscura). El tamaño del horizonte de eventos ℓ_{HE}, la frontera entre la parte influenciable y la parte no influenciable, es igual a la anchura del cono de luz futuro y decrece con el tiempo (linea interrumpida). Si la expansión es suficientemente lenta, el modelo no tiene horizonte de eventos, ya que el cono de luz intersecta las lineas de universo de todas las galaxias (derecha). Las influencias causales llegan por lo tanto arbitrariamente lejos.*

geométrica entre dos puntos viene dada por $D(t) = A_0 t^n r$, mientras que la distancia de cono de luz es

$$\ell \;=\; \int_{t_e}^{t_d} A_0^{-1}\, t^n dt \;\;=\;\; \frac{1}{1-n}\, A_0^{-1}\left[t_d^{1-n} \;-\; t_e^{1-n}\right]. \tag{13.46}$$

La presencia de un horizonte cósmico y de partículas también depende del modelo, ya que los integrales (13.42) y (13.43) divergen y convergen según el valor de n:

$$\ell_{HP} \;=\; \int_t^0 A_0^{-1}\, t'^n dt' \;=\; \begin{cases} \frac{1}{1-n} A_0^{-1}\, t^{1-n} & \text{para} \quad 0 < n < 1, \\ \infty & \text{para} \quad 1 < n, \end{cases}$$

$$\ell_{HE} \;=\; \int_t^0 A_0^{-1}\, t'^n dt' \;=\; \begin{cases} \infty & \text{para} \quad 0 < n < 1, \\ \frac{1}{1-n} A_0^{-1}\, t^{1-n} & \text{para} \quad 1 < n. \end{cases} \tag{13.47}$$

Es decir los modelos decelerados tienen un horizonte de partículas, pero no tienen horizonte cósmico, mientras que los acelerados justo al revés. La razón es que los modelos acelerados (los que tienen $n > 1$) empiezan su expansión muy lentamente, pero aceleran rápidamente (tanto más, cuanto mayor sea n). En esta expansión acelerada, las velocidades de receso de galaxias que inicialmente eran sublumínicas, van aumentando con el tiempo, alcanzando y sobrepasando la velocidad de la luz. En otras palabras, el radio de Hubble crece más lentamente que la expansión del universo y poco a poco las galaxias van saliendo del radio de Hubble. La luz emitida por galaxias suficientemente lejanas (es decir, con velocidades de receso suficientemente altas) nunca nos alcanzará, de modo que hay un horizonte de eventos (¡aunque éste no coincide con el radio de Hubble!). Sin embargo, debido al hecho de que la expansión inicial era lenta, no hay horizonte de partículas.

El caso decelerado es en muchos aspectos justo lo contrario: cuando $0 < n < 1$, los modelos empiezan su expansión inicial muy rápido, pero también deceleran pronto (tanto más, cuanto más baja sea n). Las galaxias que originalmente tenían una velocidad de receso superlumínica, están deceleradas hasta velocidades sublumínicas y entran en el radio de Hubble, que crece más rápido que la expansión del universo. En este caso sí es posible ver la luz emitida por galaxias

que tienen una velocidad de receso mayor que la de la luz: la velocidad de los fotones emitidos hacia nosotros es $v_{\rm rec} - 1$, que es positiva (es decir los fotones están recediendo) en las zonas de expansión superlumínica. Sin embargo, al decelerar la expansión, el radio de Hubble crece más rápido que la expansión y en cierto momento alcanza a los fotones. Una vez dentro del radio de Hubble, ya están en la región de expansión sublumínica y finalmente podrán llegar hasta nosotros. En este caso, no hay un horizonte de eventos, pero debido a la expansión rápida inicial, sí hay un horizonte de partículas.

13.5. El contenido de energía y materia del universo

Como hemos visto, la evolución del universo depende de la densidad de energía ρ y de la presión P, de modo que hay que especificar éstas para poder resolver las ecuaciones de Friedmann. Sin embargo, a su vez, la densidad de energía y la presión cambian con la evolución del universo y dependen por lo tanto del factor de escala. Necesitamos entonces información adicional, que determina como varía $a(t)$ con la densidad y la presión.

Esta información nos la dará la ley de conservación de energía,

$$\nabla_\mu T^{\mu\nu} \;=\; 0. \tag{13.48}$$

Aunque (13.48) es una ecuación vectorial, sólo la componente temporal nos proporciona una relación entre ρ, P y $a(t)$: sustituyendo (13.24) y (13.28) en la ley de conservación de energía, encontramos en coordenadas comóviles

$$\dot\rho \;+\; 3\,\frac{\dot a}{a}\,(\rho + P) \;=\; 0. \tag{13.49}$$

Esta ecuación, por rara que pueda parecer a primera vista, es en realidad una identidad conocida de la termodinámica. Multiplicando (13.49) por a^3, podemos reescribirla como

$$\frac{d}{dt}\Big[a^3\rho\Big] = -P\,\frac{d}{dt}\Big[a^3\Big]. \tag{13.50}$$

Si interpretamos $a^3(t)$ como el volumen de un trozo de la sección espacial en el momento t, vemos que la ley de conservación de energía dice que el cambio de energía en un volumen es igual a menos la presión por el cambio de volumen. En otras palabras, hemos recuperado una formulación de la primera ley de la termodinámica

$$dE = -PdV. \tag{13.51}$$

Con la ley de conservación de energía podemos demostrar que las dos ecuaciones de Friedmann en realidad no son independientes: derivando la ecuación de Friedmann (13.29a) con respecto a t y usando (13.49), obtenemos depués de un poco de cálculo la ecuación de aceleración (13.30). La ecuación de Friedmann y la conservación de energía implican por lo tanto la ecuación de aceleración y consecuentemente la de evolución. Dado que siempre trabajaremos con fluidos perfectos, que satisfacen la conservación de la energía, en la práctica sólo tenemos que resolver la ecuación de Friedmann para determinar la evolución del sistema.

La ley de conservación de energía (13.49) es imposible de resolver, si no especificamos con qué tipo de energía estamos tratando. El tipo de energía o matería viene especificado por la dependencia de la presión P_α de la densidad ρ_α, expresado en la *ecuación de estado*

$$P_\alpha \;=\; w_{(\alpha)}\,\rho_\alpha, \tag{13.52}$$

donde $w_{(\alpha)}$ es el *parámetro de la ecuación de estado*.[8] En principio $w_{(\alpha)}$ no tiene porqué ser constante, pero la homogeneidad y la isotropía de la métrica de FRW obliga a $w_{(\alpha)}$ sea independiente de las coordenadas x. Además se suele tomar $w_{(\alpha)}$ también independiente de t: cada fluido perfecto está caracterizado por un valor de $w_{(\alpha)}$ y, como veremos en breve, diluye de manera diferente con la expansión del universo. Si en distintas épocas el universo está dominado por distintos tipos de energía, es preferible caracterizar estas por varios fluidos perfectos, que por uno solo con un parámetro de estado variable en el tiempo.

Como hemos dicho ya, cada valor de $w_{(\alpha)}$ define un tipo de fluido perfecto. Por ejemplo, $w = 0$ corresponde a un fluido perfecto con solamente densidad de materia, sin presión y describe por lo tanto materia fría, sin interacciones, o polvo. Por otro lado $w = 1/3$ corresponde a materia muy caliente, materia relativista o radiación y $w = -1$ corresponde a una constante cosmológica. El hecho de que estos valores de w correspondan a estos tipo de fluidos se deriva a través de las leyes de la termodinámica, lo que nos llevaría fuera del ámbito de este curso, pero sí hay argumentos que nos dan cierta intuición de que sea así.

Sustituyendo la ecuación de estado (13.52) en la ecuación de conservación de energía (13.49), encontramos la siguiente ecuación diferencial para ρ_α en términos de a,

$$\dot{\rho}_\alpha \; + \; 3(w_{(\alpha)} + 1) \, \frac{\dot{a}}{a} \, \rho_\alpha \; = \; 0, \tag{13.53}$$

que sí podemos resolver. En general tenemos que la densidad varía con el tiempo como

$$\rho_\alpha(t) \; = \; \rho_0 \, a^{-3(w_{(\alpha)}+1)}(t), \tag{13.54}$$

donde ρ_0 es la densidad en un momento dado $t = t_0$ (por ejemplo en la actualidad) en que normalizamos el factor de escala como $a(t_0) \equiv 1$. Vemos por la tanto que cada tipo de fluido perfecto evoluciona de manera distinta en la expansión del universo. En particular, para materia fría ($w = 0$) vemos que la densidad va como a^{-3}, es decir, la materia se diluye de manera inversamente proporcional al volumen. Por otro lado, la densidad de energía de radiación ($w = 1/3$) evoluciona como a^{-4}, es decir aparte de diluirse inversamente proporcional al volumen, pierde energía en el corrimiento hacia el rojo de la radiación, ya que al expandirse el universo, también aumenta la longitud de las ondas de la radiación de manera lineal en a. Por último, una constante cosmológica ($w = -1$) proporciona una densidad de energía constante por unidad de volumen. En este sentido, una constante cosmológica realmente corresponde a la energía del vacío.

Una observación interesante surge al sustituir la ecuación de estado en la ecuación de aceleración:

$$\frac{\ddot{a}}{a} \; = \; -\frac{1}{6}\kappa \, (1 + 3w) \, \rho. \tag{13.55}$$

Una expansión acelerada del universo sólo es posible cuando el universo está dominado por un fluido perfecto con parámetro de estado $w < -1/3$.[9] Un universo con materia fría o radiación sufrirá una deceleración, debido a la atracción gravitatoria del contenido de energía y materia.

Otra observación importante sacamos de la ecuación de Friedmann: para que las secciones espaciales sean planas ($k = 0$), es preciso que la densidad de energía en el universo sea igual a una densidad crítica

$$\rho_c \; = \; \frac{3}{\kappa} \left(\frac{\dot{a}}{a}\right)^2. \tag{13.56}$$

Si la densidad es menor (mayor) que la densidad crítica, necesariamente tenemos que $k = -1$ ($k = +1$) y por lo tanto las secciones espaciales tienen necesariamente que tener curvatura negativa (positiva). En este caso se suele hablar de un universo abierto (cerrado), mientras que $k = 0$

[8] Anotamos $w_{(\alpha)}$ con el índice entre paréntesis, para enfatizar que en la ecuación de estado no estamos asumiendo una sumación sobre α.

[9] Siempre y cuando la densidad de energía es positiva, como en la gran mayoria de los fluidos realistas. Para una constante cosmológica negativa sin embargo, las conclusiones son justo al revés.

se le denomina un universo plano (¡refiriéndose obviamente a las secciones espaciales, no a la curvatura cuadrimensional!).

Podemos definir el *parámetro de densidad total*, que mide la densidad total en términos de la dénsidad crítica

$$\Omega_{\rm tot} \ = \ \frac{\rho}{\rho_c}. \tag{13.57}$$

Como hemos dicho antes, el universo es espacialmente plano para $\Omega_{\rm tot} = 1$, abierto para $\Omega_{\rm tot} < 1$ y cerrado para $\Omega_{\rm tot} > 1$. También es conveniente definir los *parámetros de densidad parciales* de cada componente del fluido perfecto

$$\Omega_\alpha \ = \ \frac{\rho_\alpha}{\rho_c}, \tag{13.58}$$

que mide la importancia de cada fluido en comparación con la densidad crítica. Por construcción tenemos que

$$\sum_\alpha \Omega_\alpha \ = \ \Omega_{\rm tot}. \tag{13.59}$$

13.6. Soluciones cosmológicas

Dedicaremos el resto de este capítulo a la derivación y discusión de algunas soluciones cosmológicas concretas. Aunque ahora creemos que ninguna de estas soluciones describe bien nuestro universo, merece la pena dedicarles un poco de atención por su interés histórico y matemático.

El universo estático de Einstein

En 1917 Einstein fue el primero en aplicar sus ecuaciones al universo entero y presentar un modelo cosmológico. Curiosamente su *universo estático*, no sólo fue la primera solución cosmológica, sino también, junto con el universo de Einstein-De Sitter, una de las pocas soluciones exactas que Einstein mismo obtuvo de su ecuación.

Guiado por los prejuicios científicos de la época, que requerían que el universo fuera estático, y obligado por las ecuaciones de Friedmann, que dejan bien claro que un universo dominado por materia normal (materia fría o radiación, ...) necesariamente se está expandiendo o contrayendo, Einstein se vio forzado a argumentar *"que las ecuaciones de gravedad que he defendido hasta ahora necesitan una pequeña modificación."* Para conseguir una solución cosmológica estática, tenía que introducir la constante cosmológica Λ, con dimensiones de $\mathbf{ML}^{-3}$, en las ecuaciones de Einstein,

$$R_{\mu\nu} \ - \ \frac{1}{2} g_{\mu\nu} R \ + \ \kappa\Lambda\, g_{\mu\nu} \ = \ -\kappa T_{\mu\nu}, \tag{13.60}$$

que representa la densidad de energía del vacío, $\rho_\Lambda = \Lambda$ (un fluido perfecto con $w_\Lambda = -1$) y genera un tipo de fuerza cósmica repulsiva si Λ es positivo y atractiva si Λ es negativo. Curiosamente, añadir una constante cosmológica es la única manera en que se puede generalizar la acción de Einstein-Hilbert

$$S = \int d^4x \, \sqrt{|g|} \, \Big[\tfrac{1}{2\kappa} R \ - \ \Lambda \Big], \tag{13.61}$$

sin introducir extra grados de libertad, ni romper la covariancia general, ni generar ecuaciones diferenciales de orden más alto que 2. En 1915 Einstein no incluyó este término, porque no tenía ninguna razón para hacerlo, pero desde su introducción en 1917, la constante cosmológica ha representado uno de los desafíos más grandes de la física, como veremos en el siguiente capítulo.

Si insistimos en un universo estático, es decir $\dot{a}(t) = \ddot{a}(t) = 0$, la ecuación de aceleración,

$$0 \ = \ -\frac{1}{6}\kappa \ \sum_\alpha (1 \ + \ 3w_{(\alpha)}) \ \rho_\alpha, \tag{13.62}$$

nos obliga a tener aparte de las formas normales de energía, un fluido perfecto con un parámetro de estado suficientemente negativo para compensar la atracción del resto de energía. Históricamente Einstein derivó su solución para un universo con materia fría y una constante cosmológica, pero se puede sacar también una solución análoga con radiación en vez de materia fría. Tomando en cuenta que para la materia fría $w_M = 0$, vemos que la ecuación de aceleración está satisfecha si la densidad de materia y la densidad de energía del vacío están relacionadas como

$$\rho_M \;=\; 2\rho_\Lambda. \tag{13.63}$$

No sólo implica esto que la constante cosmológica tiene que ser positiva (ya que $\rho_M > 0$), sino también sugiere un primer problema con la solución: para que la solución se mantenga estática, es preciso que las densidades de los dos fluidos estén exactamente ajustadas según la relación (13.63). En realidad no hay ninguna razón a priori para suponer que las densidades se comportan como exige la solución, ya que no existe ningún mecanismo dinámico que lleve a ello. Por lo tanto, por muy imprecindible que sea para la existencia de esta solución, la relación (13.63) parece completamente ad hoc. Esto es lo que se suele llamar en la física un *problema de ajuste fino*, y suele levantar sospecha en los casos donde aparece.

Aun así, tomando en cuenta (13.63), la ecuación de Friedmann coge la forma

$$0 \;=\; \frac{1}{2}\kappa\,\rho_M \;-\; a^{-2}k, \tag{13.64}$$

y dado que $\rho_M > 0$, tenemos necesariamente que $k = 1$. En otras palabras, se puede representar el universo estático de Einstein,

$$ds^2 \;=\; dt^2 \;-\; \frac{2}{\kappa\rho_M}\Big[d\chi^2 \;+\; \sin^2\chi\; d\Omega_2^2\Big], \tag{13.65}$$

como un cilindro, donde el eje del cilindro es el tiempo y las secciones ortogonales al eje representan las secciones espaciales $\mathbb{S}^3$ (véase Figura 13.5). La luz da la vuelta al universo en un tiempo finito y cada observador puede ver la parte atrás de su cabeza. Objetos en las antípodas del universo parecen muy cerca, ya que los conos de luz de un evento primero divergen, pero luego vuelven a converger, para coincidir todos en el mismo momento en el punto opuesto del universo al que han salido.

En 1930 Eddington demostró que el universo estático de Einstein no es estable, por la razón que ya mencionamos antes: necesita un delicado equilibrio entre la atracción gravitatoria de la materia y la fuerza repulsiva de la constante cosmológica y cualquier perturbación que rompa este equilibrio hará que predomine una de las dos fuerzas y causará la expansión o el colapso del universo. Pero el verdadero golpe de gracia para este modelo fue el descubrimiento de Hubble en 1929 de que el universo no es estático, sino que las galaxias muestran un corrimiento hacia el rojo, debido a la expansión del universo. Einstein retiró su modelo, refiriéndose a la constante cosmológica como el error más grande de su vida. Curiosamente, desde 1998 hay indicios observacionales muy fuertes (suficientemente fuertes para otorgar el Premio Nobel de Física del 2011 a los astrónomos Saul Perlmutter (1959 -), Brian Schmidt (1962 -) y Adam Riess (1969 -), por el descubrimiento), de que el universo está actualmente en una fase de expansión acelerada, de manera muy parecida a como si hubiera una constante cosmológica positiva.

El espacio de De Sitter

En el mismo año que Einstein presentó su universo estático, el astrónomo holandés Willem de Sitter (1872 - 1934) sorprendió a la comunidad científica con una solución de un universo vacío (en el sentido de que no tiene ni materia, ni radiación), pero en aceleración constante, debido a la presencia de una constante cosmológica positiva.

Concretamente el espacio de De Sitter tiene $\rho_M = \rho_{\rm rad} = 0$ y $\rho_\Lambda > 0$ con $k = 0$. Con estos parámetros se puede integrar la ecuación de Friedmann directamente, obteniendo como solución

$$a(t) = e^{t/R_0}, \tag{13.66}$$

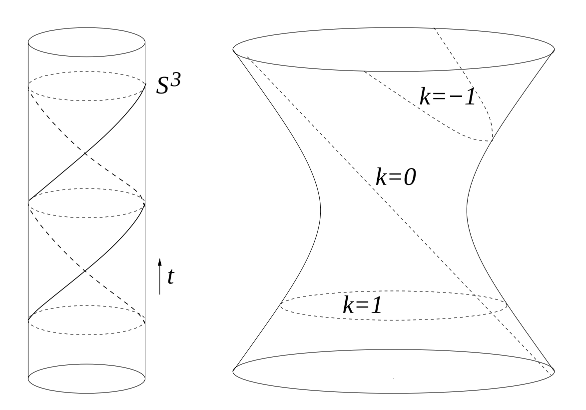


Figura 13.5: *El universo estático de Einstein (izquierda) y el espacio de De Sitter (derecha): En el universo estático de Einstein, las secciones espaciales son tres-esferas $\mathbb{S}^3$ y la luz da la vuelta al universo en un tiempo finito. El espacio de De Sitter se puede describir en coordenadas FRW tanto con secciones espaciales planas ($k = 0$), esféricas ($k = 1$) o hiperboloides ($k = -1$), dependiendo del sistema de coordenadas que se elija.*

donde $R_0 = \sqrt{3/\kappa\Lambda}$ es el *radio de De Sitter*, que es una medida para la curvatura de este espacio. El espacio de De Sitter,

$$ds^2 \;=\; dt^2 \;-\; e^{2t/R_0}\,\delta_{ij}\,dx^i dx^j, \tag{13.67}$$

representa un universo vacío en expansión exponencial (el parámetro de deceleración (13.44) vale $q = -1$), por causa de la fuerza repulsiva de la constante cosmológica. Dos observadores, que inicialmente están arbitrariamente cercanos, se alejarán cada vez más, hasta perderse literalmente de vista. Llegará un momento en que la expansión es tan grande, que a la luz no le da tiempo de llegar de un observador a otro y los observadores ya no tendrán manera de comunicarse. Efectivamente no es difícil ver que el espacio de De Sitter tiene un horizonte cósmico a una distancia $\ell_{HE} = R_0 e^{-t/R_0}$ del observador. Es decir la parte del espacio que un observador puede llegar a influenciar decrece exponencialmente con el tiempo.

Sin embargo, a la vez es un espacio donde nada cambia: la aceleración es constante a lo largo del tiempo y no tiene materia o radiación que pueda diluirse, de modo que en cada instante el espacio de De Sitter es un rescaleo de los momentos anteriores.[10] Igual que el universo estático, el espacio de De Sitter no tiene principio, ni fin, ha existido siempre y seguirá existiendo eternamente.

Es interesante calcular el tensor de Riemann del espacio de De Sitter:

$$R_{titj} \;=\; -R_0^{-2}\, e^{2t/R_0}\, \tilde{g}_{ij}, \qquad\qquad R_{ijkl} \;=\; -R_0^{-2}\, e^{4t/R_0}\, (\tilde{g}_{il}\tilde{g}_{jk} - \tilde{g}_{ik}\tilde{g}_{jl}), \tag{13.68}$$

puesto que satisface la condición (13.1) de una métrica lorentziana con curvatura constante positiva con radio R_0. Efectivamente, al ser máximamente simétrico, el espacio de De Sitter es el análogo cuadrimensional y lorentziano de una esfera. Podemos por lo tanto pensar en este espacio como la solución del vacío de la acción (13.61) en presencia de una constante cosmológica positiva, igual que el espacio de Minkowski es el vacío de la acción de Einstein-Hilbert.

Uno se podría preguntar cuál sería la solución correspondiente a un universo vacío con constante cosmológica positiva, pero con secciones espaciales con $k = \pm 1$. Un cálculo bastante directo

[10] En el lenguaje de los Capítulos **??** y **??**: ∂_t no es un vector de Killing, pero sí un vector de Killing conforme.

muestra que estas soluciones vienen dadas por

$$ds^2 = d\tilde{t}^2 - R_0^2 \cosh^2(R_0^{-1}\tilde{t}) \Big[d\chi^2 + \sin^2\chi \, d\Omega_2^2\Big], \qquad (k=+1),$$

$$ds^2 = d\bar{t}^2 - R_0^2 \sinh^2(R_0^{-1}\bar{t}) \Big[d\chi^2 + \sinh^2\chi \, d\Omega_2^2\Big], \qquad (k=-1). \tag{13.69}$$

Ahora, al calcular el tensor de Riemann de estas métricas, encontramos que también satisfacen la condición (13.1) de curvatura constante positiva con radio R_0. Puesto que la solución de la ecuación (13.1) es única para cada valor de K (por lo menos localmente), sabemos que en realidad las métricas (13.69) también describen el mismo espacio de De Sitter, pero en coordenadas diferentes. De hecho, aunque t, $\tilde{t}$ y $\bar{t}$ jueguen cada uno el papel de coordenada temporal en la métrica de Friedmann-Robertson-Walker (13.22), son en realidad distintas maneras de parametrizar el tiempo (por eso sus nombres distintos). Lo que acabamos de encontrar es que es posible foliar el espacio de De Sitter de varias maneras, tanto en secciones espaciales abiertas, cerradas y planas, gracias a la cantidad de simetría. En este caso, la geometría de las secciones espaciales es claramente una propiedad que depende del sistema de coordenadas elegido. Entreremos en más detalle sobre las distintas formas de la métrica de De Sitter en la sección **??**.

La paradoja de los modelos cosmológicos de 1917 por lo tanto era que había dos soluciónes: una de Einstein, con materia, pero sin movimiento, y una de De Sitter, con movimiento, pero sin materia. Sin embargo estas dos soluciones no tardaron a insprirar otra gente, como Friedmann y Lemaître a considerar modelos más realistas.

El espacio de anti-De Sitter

Aunque nunca fue considerado como un modelo cosmológico realista, es útil preguntar cuál sería el vacío de la acción (13.61) con constante cosmológica $\Lambda < 0$, es decir $\rho_M = \rho_{\rm rad} = 0$ y $\rho_\Lambda < 0$. Un vistazo a las ecuaciones de Friedmann deja bien claro que tal solución sólo es posible en coordenadas FRW para $k = -1$.

Un cálculo muy parecido al caso anterior muestra que la solución, llamado el *espacio de anti-De Sitter* es de la forma

$$ds^2 = dt^2 - R_0^2 \sin^2(R_0^{-1}t) \Big[d\chi^2 + \sinh^2\chi \, d\Omega_2^2\Big], \tag{13.70}$$

donde ahora $R_0 = \sqrt{-3/\kappa\Lambda}$. El espacio de anti-De Sitter en cierto modo es lo contrario del espacio de De Sitter: en vez de ser sometido una expansión exponencial, sufre una contracción continua. Partículas emitidas desde un punto inicialmente divergen, pero la atracción de la constante cosmológica negativa las frena y las hacen recolapsar otra vez en un tiempo finito, como indica el factor de escala.

Sin embargo, el espacio de anti-De Sitter mismo no colapsa: igual que el espacio de De Sitter es perfectamente regular en todo momento (en realidad $t = 0$ y $t = R_0\pi$ son singularidades de coordenadas), e incluso satisface la condición (13.1) con $K = R_0^{-2}$ (ejerc.) En otras palabras, anti-De Sitter es un espacio de curvatura constante negativa, el análogo lorentziano cuadrimesional del hiperboloide y por lo tanto podemos identificarlo como el vacío de (13.61) con $\Lambda < 0$.

Cabe preguntar entonces, si anti-De Sitter también es máximamente simétrico, igual que De Sitter, ¿por qué sólo se puede foliar con secciones espaciales abiertas? La respuesta es que también existen foliaciones del espacio de anti-De Sitter con geometrías planas y curvatura positiva, pero la métrica con estas foliaciones no son del tipo FRW. (La diferencia está en que la coordenada perpendicular a las hipersuperficies de la foliación es una coordenada espacial). El espacio de anti-De Sitter no se suele tomar en serio como modelo cosmológico (menos aún ahora que hay indicaciones obervacionales de que el universo pueda tener una constante cosmológica positiva), pero sí se usa mucho como herramienta matemática en otros terrenos de la física teórica, especialmente en teorías de holografía (véase sección 11.6). Discutiremos más extensamente las

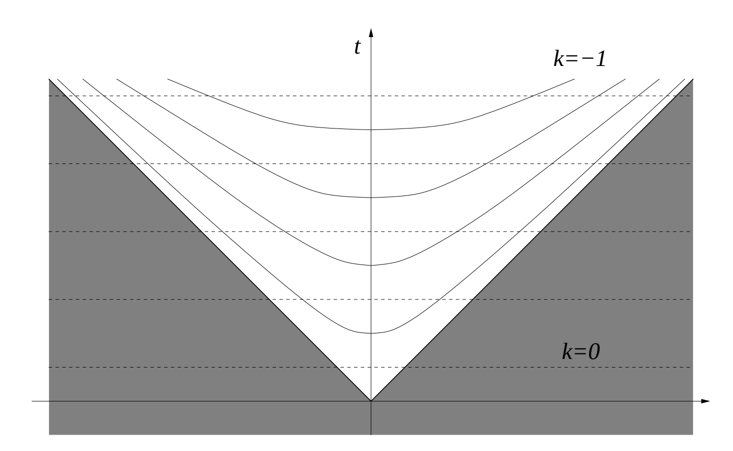


Figura 13.6: *El universo de Milne como una sección del espacio de Minkowski: el universo de Milne es la parte del espacio de Minkowski contenida dentro del cono de luz futuro del origen (parte en blanco). Las seccíónes espaciales con $k = -1$ en las coordenadas FRW son superficies de puntos a la misma distancia temporal del origen (lineas continuas). Las secciones planas ($k = 0$, lineas discontinuas) de las habituales coordenadas cartesianas sí cubren todo el espacio. Obsérvese que la coordenada t del universo de Milne ($k = -1$) no es la misma coordenada que la coordenada temporal en cartesianas.*

propiedades geométricas y las formas de la métrica de anti-De Sitter en la sección **??**.

El universo de Milne

El *universo de Milne* es una solución del las ecuaciones del vacío, en ausencia de materia, de radiación, e incluso de una constante cosmológica. En otras palabras, $\rho_M = \rho_{\rm rad} = \rho_\Lambda = 0$. En este caso, las ecuaciones de Friedmann,

$$\left(\frac{\dot a}{a}\right)^2 = -a^{-2}k, \qquad \frac{\ddot a}{a} = 0, \tag{13.71}$$

dejan dos posibilidades: o bien $k = 0$, ó bien $k = -1$. El primer caso es trivial, ya que sólo permite la solución $a(t) = a_0$, una constante, y la solución es fácilmente identificable como el espacio de Minkowski.

El otro caso, $k = -1$, el universo de Milne, tiene como solución $a(t) = t$ (más una constante de integración, que se puede absorber en una redefinición de la coordenada t), de modo que la métrica del universo de Milne coge la forma

$$ds^2 = dt^2 - t^2 \Big[d\chi^2 + \sinh^2\chi \, d\Omega_2^2\Big]. \tag{13.72}$$

A primera vista parece que la solución es singular para $t \to 0$. Sin embargo, al calcular el tensor de Riemann para obtener algún invariante de curvatura, vemos que

$$R_{titj} = R_{ijkl} = 0. \tag{13.73}$$

En otras palabras, también para el caso $k = -1$, ¡el universo de Milne no es más que el espacio de Minkowski! Para ser más exactos: es un trozo de él: la métrica (13.72) representa la sección de Minkowski que está contenida dentro del cono de luz futuro del origen de las coordenadas $(t, \chi, \theta, \varphi)$ (véase Figura 13.6). Las secciones espaciales son los conjuntos de puntos a la misma distancia temporal t del origen y por la signatura lorentziana del espacio de Minkowski tienen curvatura constante negativa (obsérvese que la coordenada temporal t en la métrica (13.72) no

coincide con la coordendas temporal de la base cartesiana). En el límite $t \to 0$ las secciones espaciales se aproximan al cono de luz, que es una superficie nula, lo que resulta en una singularidad de coordenadas. Sin embargo, como sabemos, el espacio de Minkowski es perfectamente regular en todos los puntos.

El universo de Milne fue presentado en 1948 por el astrofísico británico Edward Arthur Milne (1896 - 1950), no tanto para presentar una solución cosmológica, sino en un intento de buscar una nueva teoría para la gravedad, una alternativa a la relatividad general, llamada *Relatividad Cinemática*, una teoría que, aparte de Milne mismo, nadie nunca ha tomado en serio.

Quizá la gran lección del universo de Milne es que uno siempre debería ser crítico con las soluciones que uno obtiene. No siempre es fácil darse cuenta si una solución nueva no es más que una solución conocida en otras coordenadas. En particular, se suele decir que el espacio de Minkowski es la solución que más veces ha sido redescubierta...

El espacio de Einstein-De Sitter y otros universos espacialmente planos

El espacio de Einstein-De Sitter es el primer modelo más o menos realista que encontramos en este repaso de soluciones históricas. De hecho, aunque presentado en 1932 en un artículo conjunto de los dos científicos que aparecen en su nombre, fue considerado una buena descripción de nuestro universo hasta bien avanzado los años 1980.

El espacio de Einstein-De Sitter supone que el universo consiste principalmente de materia fría con densidad crítica, o sea $\Omega_M = 1$ y $\Omega_{\text{rad}} = \Omega_\Lambda = 0$. Debido a la densidad crítica de materia, las secciones espaciales son planas ($k = 0$) y la ecuación de Friedmann,

$$\Big(\frac{\dot{a}}{a}\Big)^2 \;=\; \frac{1}{3}\kappa\rho_M, \tag{13.74}$$

es directamente integrable, si recordamos que la materia fría se diluye con el factor de escala como $\rho_M(t) = \rho_0\, a^{-3}(t)$. La solución a esta ecuación diferencial es

$$a(t) = \Big(\frac{3}{2}H_0 t\Big)^{\frac{2}{3}}, \tag{13.75}$$

donde la constante $H_0 = \sqrt{\frac{1}{3}\kappa\rho_0}$ es el parámetro de Hubble en el momento $t = t_0$ (cuando $\rho(t_0) = \rho_0$ y $a(t_0) = 1$). La métrica del espacio de Einstein-De Sitter es por lo tanto de la forma

$$ds^2 \;=\; dt^2 \;-\; \Big(\frac{3}{2}H_0 t\Big)^{\frac{4}{3}}\delta_{ij}dx^i dx^j, \tag{13.76}$$

y describe un universo en expansión decelerada ($q = 1/2$). Esta métrica parece singular para $t \to 0$ y efectivamente el escalar de Ricci,

$$R \;=\; \frac{4}{3}\, t^{-2}, \tag{13.77}$$

diverge para $t = 0$, lo que implica que la singularidad es física. La coordenada temporal corre por lo tanto en el intervalo $]0, \infty[$, donde la singularidad en $t = 0$ representa el origen del universo. Se suele llamar *Big Bang* tanto a la singularidad inicial, como a los modelos cosmológicos que la predicen. El nombre ("Gran Bum") lo puso el astrofísico británico Fred Hoyle (1915 - 2001) en una entrevista con la BBC en 1950, para mofarse de la idea, puesto que él era partidario de la teoría del estado estacionario, que predice un universo sin evolución, de edad infinita. El espacio de Einstein-De Sitter es por lo tanto el primer modelo que encontramos según el cual el universo no ha existido desde siempre.[11]

[11] Hay que tener un poco de cautela con esta última frase: Con "desde siempre" queremos decir un tiempo infinitamente

La edad del universo se calcula fácilmente de la expresión (13.75) del factor de escala: denominando la actualidad $t = t_0$ y normalizando el factor de escala como $a(t_0) \equiv 1$, vemos que la edad del universo en el modelo de Einstein-De Sitter viene dada por

$$t_0 = \frac{2}{3} H_0^{-1}. \tag{13.78}$$

Hasta la segunda mitad de los años 1990, se estimaba el valor actual del parámetro de Hubble H_0 entre 50 y 90 km/s/Mpc,[12] lo que resultaba en una edad entre aproximadamente 6,5 y 13 mil millones de años. La cota superior en realidad está cerca del valor que se cree hoy en día, pero la cota inferior es claramente demasiado corta, puesto que hay estrellas y estructuras que son más viejos que esto. El problema más grave por lo tanto del modelo de Einstein y De Sitter era que predecía una edad demasiado corta para nuestro universo. En el siguiente capítulo veremos como la cosmología moderna resuelve este problema.

Merece la pena mirar la evolución de universos espacialmente planos dominados por otro tipo de energía, es decir para w general. Tomando en cuenta que la densidad de energía en general desminuye como $\rho(t) = \rho_0\, a^{-3(w+1)}$, la ecuación de Friedmann se reduce a

$$\dot{a}^2 \;=\; H_0^2\, a^{-3(w+1/3)}, \tag{13.79}$$

con H_0 definido como arriba. La solución general viene dada por

$$a(t) \;=\; \Big[\frac{3(1+w)}{2}\; H_0\, t\Big]^{\frac{2}{3(1+w)}}, \tag{13.80}$$

siempre y cuando $w \neq -1$. El caso $w = -1$ requiere un análisis propio, pero en realidad este caso es él del espacio de De Sitter con secciones planas (13.67).

Obsérvese que el factor de escala depende del parámetro de estado w, y por lo tanto universos con distintos contenidos de energía, tienen evoluciones diferentes: en el espacio de Einstein-De Sitter el factor de escala va como $a(t) \sim t^{\frac{2}{3}}$, pero por ejemplo un universo dominado por radiación ($w = 1/3$) crece como $a(t) \sim t^{\frac{1}{2}}$. Es decir, en tiempos primodiales, un universo con radiación se expande más rápido que un universo con materia, pero también decelera más rápido, de modo que tarde o temprano su factor de escala será alcanzado por el factor de escala de Einstein-De Sitter y quedará atrás (véase Figura 13.7). Efectivamente, un universo dominado por radiación tiene $q = 1$, el doble que el universo de Einstein-De Sitter. En general el parámetro de deceleración de estos modelos es $q = \frac{1}{2}(1 + 3w)$: En otras palabras, los universos con $w > -1/3$ tendrán una expansión decelerada, mientras los con $w < -1/3$ serán acelerados. Nótese que esto coincide con el análisis que hemos hecho en la sección 13.4. Sabemos por lo tanto que los primeros tienen un horizonte de partículas, mientras los últimos uno de eventos.

Si los distintos modelos evolucionan a ritmos diferentes, también quiere decir que tienen edades distintas, por lo menos si todos coinciden en el valor actual del parámetro de Hubble H_0 (como en la Figura 13.7). Igual que en el caso de Einstein-De Sitter, se calcula la edad desde el factor de escala. De (13.80) vemos que para w general (pero $w \neq -1$)

$$t_0 \;=\; \frac{2}{3(w+1)H_0}, \tag{13.81}$$

grande. Alrededor del año 400 A.D., el teólogo cristiano (San) Agustín de Hipona (354 - 430) se preguntó por qué Dios había esperado tanto en crear el mundo. Si bien la creación tuvo lugar hace 6.000 años (según los creacionistas) o hace $13 \cdot 10^9$ años (según la cosmología moderna), ¿por qué en ese momento particular y no hace 10^{90} años, hace 10^{900} años, o cualquier otra cantidad arbitrariamente grande? Según Agustín, en realidad la pregunta no tiene sentido, porque durante la creación Dios no sólo creó el mundo (el espacio), sino también a la vez el tiempo y por lo tanto no hay un "antes de la creación". Y si Dios es eterno, afirma Agustín, es porque Él mismo está fuera del tiempo (y del espacio). Incluso sin juzgar sus ideas metafísicas, hay que admitir que el concepto que tiene Agustín sobre el espacio y el tiempo es sorprendentemente relativista, para ser del siglo IV A.D.

[12]El parámetro de Hubble se suele medir en kilómetros por segundo por Megapársec (1 Mpc son aproximadamente $3,3 \cdot 10^6$ años-luz).

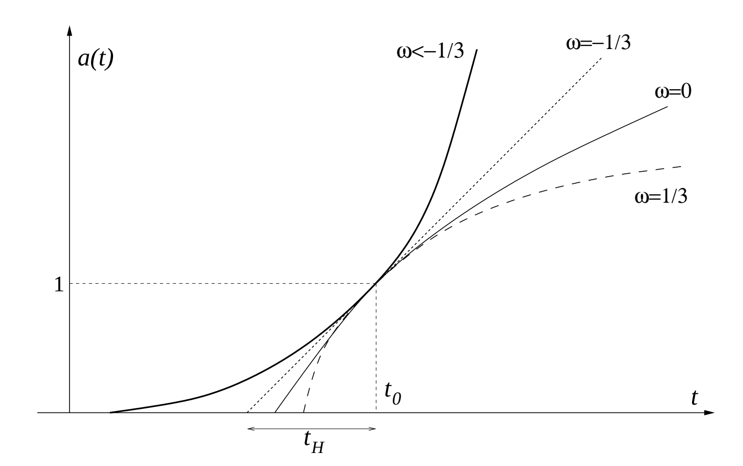


Figura 13.7: *La evolución del factor de escala para distintos valores de w (para universos espacialmente planos): un universo dominado por materia fría (linea fina continua) inicialmente crece más lentamente que un universo dominado por radiación (linea discontinua), pero decelera menos rápido, de modo que después de un cierto tiempo crecerá más rápido. Todos los universos con $w > -\frac{1}{3}$ deceleran, el universo con $w = -\frac{1}{3}$ (linea punteada) crece de manera constante y universos con $w < -\frac{1}{3}$ (linea continua en negrita) son acelerados. Los universos tienen una edad más elevada, cuanto más bajo sea w. Por otro lado, puesto que todos tienen el mismo valor para H_0, el tiempo de Hubble t_H es el mismo para todos.*

y por lo tanto el universo es más joven, cuanto más alto el valor de w.

La métrica general para universos espacialmente planos (con $w \neq -1$)

$$ds^2 \;=\; dt^2 \;-\; \Big[\frac{3(1+w)}{2}\, H_0\, t\Big]^{\frac{4}{3(1+w)}} \delta_{ij} dx^i dx^j, \tag{13.82}$$

tiene una singularidad del tipo Big Bang para $t = 0$: el escalar de Ricci viene dado por

$$R \;=\; -\frac{4(w - \frac{1}{3})}{(w+1)^2}\, t^{-2}, \tag{13.83}$$

lo que efectivamente es divergente para todo w, salvo $w = \frac{1}{3}$. Para saber qué es lo que pasa en este último caso, tenemos que calcular otro escalar de curvatura. Por ejemplo, vemos que para estas métricas el invariante de curvatura

$$R_{\mu\nu} R^{\mu\nu} \;=\; \frac{16(1+3w^2)}{9(w+1)^4}\, t^{-4}, \tag{13.84}$$

es singular para cualquier valor de w, y en particular para $w = \frac{1}{3}$. Todos estos universos empezaron por lo tanto con un Big Bang, salvo $w \neq -1$, que como hemos visto ha existido desde siempre.

Universos dominados por materia con k general

Por último miraremos a los modelos para universos con materia fría, que no tengan necesariamente la densidad crítica. Ya hemos dicho en la sección 13.5 que si la densidad es menor que la crítica, las secciones espaciales tienen curvatura negativa, mientras que si la densidad es mayor, las secciones espaciales son cerradas.

Calcular la métrica de estas soluciones resulta más fácil en las coordenadas conformes (13.32). En estas coordenadas, las ecuaciones de Friedmann (13.34) para un fluido perfecto con $w = 0$ se pueden escribir como

$$3(h^2 + k) \;=\; \kappa a^{-1} \rho_0, \qquad\qquad 2h' + h^2 + k \;=\; 0, \tag{13.85}$$

donde $h = a'/a$ es el parámetro de Hubble en coordenadas conformes y la prima indica una derivación con respecto al tiempo conforme τ. La ecuación de evolución se puede integrar fácilmente para τ como función de h, resultando

$$\tau \;=\; -2\int\frac{dh}{h^2+k}, \tag{13.86}$$

que se puede resolver con técnicas estándar para todos los valores de k.

Miremos primero el caso $k = 0$, a pesar de que ya sabemos que la solución es el espacio de Einstein-De Sitter, para familizarizarnos con las coordenadas conformes. En este caso la integral (13.86) se resuelve como

$$\tau \;=\; -2\int\frac{dh}{h^2} \;=\; 2h^{-1}+c, \tag{13.87}$$

de modo que podemos escribir h, y por lo tanto a, en función de τ como

$$h \;=\; \frac{2}{\tau-c} \qquad\Rightarrow\qquad a \;=\; \alpha(\tau+c)^2, \tag{13.88}$$

donde c y α son constantes de integración. Se puede reabsorber c en una redefinición de τ, pero α será determinada por la ecuación de Friedmann, que todavía nos queda por resolver. Sustituyendo la solución para a en la primera ecuación de (13.85), vemos que el factor de escala viene dado por

$$a(\tau) \;=\; \frac{1}{4}H_0^2\,\tau^2, \tag{13.89}$$

y la métrica es de la forma

$$ds^2 \;=\; \frac{1}{16}H_0^4\,\tau^4\Big[d\tau^2-\delta_{ij}dx^idx^j\Big]. \tag{13.90}$$

Para ver que esta solución es realmente el espacio de Einstein-De Sitter (13.76), sólo hace falta ir a coordenadas comóviles a través del cambio de coordenadas (13.31)

$$t \;=\; \frac{1}{12}H_0^2\,\tau^3 \qquad\Leftrightarrow\qquad \tau \;=\; \Big[12H_0^{-2}\,t\Big]^{\frac{1}{3}}. \tag{13.91}$$

Efectivamente sustituyendo la expresión para $\tau(t)$ en (13.89), obtenemos la expresión (13.75).

El caso $k = -1$ es nuevo, pero la integral (13.86) es fácilmente obtenible:

$$\tau \;=\; -2\int\frac{dh}{h^2-1} \;=\; 2\text{Arccotgh}\,h, \tag{13.92}$$

de modo que

$$h \;=\; \text{cotgh}\frac{\tau}{2} \qquad\Rightarrow\qquad a \;=\; R_0\sinh^2\frac{\tau}{2}. \tag{13.93}$$

Otra vez la ecuación de Friedmann determina la constante de integración como $R_0 = \frac{1}{3}\kappa\rho_0$, un parámetro con dimensión $\mathbf{L}$, como era de esperar,[13] de modo que métrica viene dada por

$$ds^2 \;=\; R_0^2\sinh^4\frac{\tau}{2}\Big[d\tau^2-d\chi^2-\sinh^2\chi d\Omega_2^2\Big]. \tag{13.94}$$

Al intentar escribir esta métrica en coordenadas comóviles, nos encontramos con un problema técnico. El cambio de coordenadas (13.31) puede expresar t en función de τ

$$t \;=\; \frac{1}{2}R_0\,\Big[\sinh\tau-\tau\Big], \tag{13.95}$$

[13] Obsérvese que en contraste con el caso $k = 0$, la densidad ρ_0 en un momento $t = t_0$ se mide en kilogramos por ángulo sólido (tridimensional) y tiene por lo tanto dimensión $\mathbf{M}$.

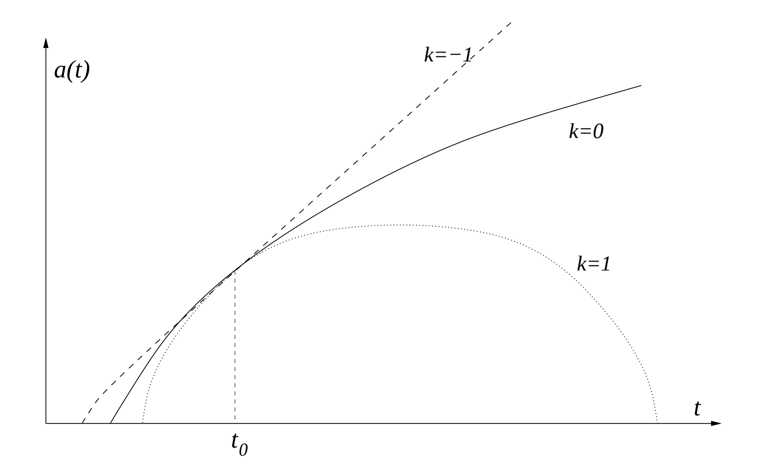


Figura 13.8: *El factor de escala para universos dominados por materia con distintos valores de k: un universo con densidad crítica ($k = 0$, linea continua) corresponde al espacio de Einstein-De Sitter; el caso subcrítico ($k = -1$, linea discontinua) expande más rápido y se aproxima en tiempos tardíos al espacio de Minkowksi; el caso supercrítico ($k = 1$, linea punteada) expande muy rápido inicialmente, pero está frenado por la gran cantidad de materia y recolapsará en un tiempo finito. Hace falta interpretar este diagrama con cuidado, puesto que $a(t)$ en el caso $k = 0$ es adimensional, mientras que tiene dimensión $\mathbf{L}$ para $k = \pm 1$.*

pero no al revés, de modo que no podemos obtener una expresión exacta para $a(t)$. Sin embargo, sí podemos aprender algunas cosas de esta métrica, mirando a los tensores de curvatura y la evolución en tiempos muy tardíos.

El escalar de Ricci viene dado por

$$R \;=\; 3R_0^{-2}\,\sinh^{-6}\frac{\tau}{2}, \qquad (13.96)$$

de modo que, al igual que el espacio de Einstein-De Sitter, tiene una singularidad del tipo Big Bang en $\tau = 0$. Sin embargo, la expansión es mucho más rápida que en el caso $k = 0$. Aunque no podemos invertir la relación (13.95) en general, sí podemos mirar lo que pasa cuando $\tau \gg 1$. En este caso tenemos que

$$t \;\approx\; \frac{1}{4}R_0\, e^{\tau}, \qquad (13.97)$$

y por lo tanto

$$a(t) \;\approx\; t. \qquad (13.98)$$

Por lo tanto vemos que un universo dominado por materia con densidad más baja que la densidad crítica se aproxima en tiempos tardíos (o no tan tardíos, ya que va exponencialmente) al universo de Milne (13.72), que a su vez habíamos identificado con el espacio de Minkowski. En otras palabras, un universo con densidad subcrítica se expande tan rápido y diluye la materia tanto, que el universo se aproxima al espacio plano. La evolución del factor de escala para todo τ está dibujada en Figura 13.8.

Por último, miremos el caso $k = 1$. Entonces la integral (13.86) se resuelve como

$$\tau \;=\; -2\int\frac{dh}{h^2+1} \;=\; 2\mathrm{Arccotg}\, h, \qquad (13.99)$$

de modo que podemos escribir h y por lo tanto a en función de τ como

$$h \;=\; \mathrm{cotg}\,\frac{\tau}{2} \qquad \Rightarrow \qquad a \;=\; R_0 \sin^2\frac{\tau}{2}. \qquad (13.100)$$

Otra vez la ecuación de Friedmann determina la constante de integración en función de la densidad ρ_0 como $R_0 = \frac{1}{3}\kappa\rho_0$, de modo que métrica viene dada por

$$ds^2 \;=\; R_0^2 \sin^4 \frac{\tau}{2} \Big[d\tau^2 - d\chi^2 - \sin^2 \chi d\Omega_2^2 \Big]. \tag{13.101}$$

También aquí encontramos problemas al intentar escribirla en coordenadas comóviles. El cambio de coordenadas (13.31) puede expresar t en función de τ

$$t \;=\; \frac{1}{2} R_0 \Big[\tau - \sin\tau \Big], \tag{13.102}$$

pero no al revés. Sin embargo, con las dos expresiones para $a(\tau)$ y $t(\tau)$ sí tenemos una parametrización de $a(t)$, que nos permite interpretar la solución: la curva $a(t)$ representa una cicloide, la curva trazada por un punto en el borde de un círculo, que rueda sin deslizarse (véase Figura 13.8). El factor de escala por lo tanto crece inicialmente muy rápido, pero está frenado por la gran cantidad de materia (recuerda que estamos en el caso super-crítico), hasta parar del todo y contraerse otra vez. Al calcular el escalar de Ricci,

$$R \;=\; 3R_0^{-2} \sin^{-6} \frac{\tau}{2}, \tag{13.103}$$

vemos que la métrica no sólo es singular en $\tau = 0$ ($t = 0$, el Big Bang), sino también en $\tau = 2\pi$, o equivalentemente en $t = \pi R_0$. El recolapso causará un aumento de la densidad que resultará en otra singularidad, parecida al Big Bang, denominado *Big Crunch* (Gran Recolapso), que hará desaparecer el universo entero. Lo curioso es que cuanto más grande la densidad del universo, más tardará en colapsarse.

Vemos por lo tanto que el espacio de Einstein-De Sitter es justo el caso límite entre un universo cerrado y uno abierto. Teniendo justo la densidad crítica, no hay suficiente materia para frenar la expansión por completo y iniciar un recolapso, ni tampoco tan poca que ésta se diluya del todo y haga que el universo se convierte en Minkowski.

En este capítulo hemos repasado algunos de los modelos cosmológicos históricos, que en su día fueron considerados atractivos por razones que hoy en día nos pueden parecer poco relevantes o incluso ingenuas. En el siguiente capítulo daremos una descripción más realista de lo que se cree que es nuestro universo.

Apéndice A

Convenios

Es más probable que dos físicos compartan el mismo cepillo de dientes
que los mismos convenios.
(Anónimo)

En la mayor parte del libro utilizamos el convenio de sumación de Einstein, donde se supone que se suma sobre índices superiores e inferiores repetidos:

$$V_\mu W^\mu \equiv \sum_{\mu=1}^{N} V_\mu W^\mu. \tag{A.1}$$

Además trabajamos en unidades naturales, donde $c = 1$.

A.1. Relatividad especial

Un punto en el espacio de Minkowski está caracterizado por un vector cuadrimensional x^μ, con componentes

$$x^\mu = \begin{pmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \end{pmatrix} = \begin{pmatrix} t \\ x \\ y \\ z \end{pmatrix}. \tag{A.2}$$

La métrica de Minkowski en coordenadas cartesianas viene dada por

$$\eta_{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} = \text{diag}\,(1,-1,-1,-1), \tag{A.3}$$

de modo que el elemento de linea es de la forma

$$ds^2 \;=\; \eta_{\mu\nu} dx^\mu dx^\nu \;=\; dt^2 - dx^2 - dy^2 - dz^2. \tag{A.4}$$

Una transformación de Lorentz $\Lambda^\mu{}_\nu$ es una transformación lineal en el espacio de Minkowski que deja invariante la métrica

$$\eta_{\mu\nu} = \Lambda^\rho{}_\mu \Lambda^\lambda{}_\nu \eta_{\rho\lambda}. \tag{A.5}$$

Un vector contravariante V^μ y un vector covariante V_μ son objetos de 4 componentes que transforman bajo una transformación de Lorentz $\Lambda^\mu{}_\nu$ como

$$V'^\mu = \Lambda^\mu{}_\nu V^\nu, \qquad\qquad V'_\mu = (\Lambda^{-1})^\nu{}_\mu V_\nu. \tag{A.6}$$

Contracción con la métrica $\eta_{\mu\nu}$ convierte un vector covariante en contravariante y vice versa,

$$V_\mu = \eta_{\mu\nu} V^\nu, \qquad\qquad V^\mu = \eta^{\mu\nu} V_\nu. \tag{A.7}$$

Un tensor de rango (m, n) $T^{\mu_1 \ldots \mu_m}{}_{\nu_1 \ldots \nu_n}$ es un objeto de 4^{m+n} componentes que transforma como

$$T'^{\mu_1 \ldots \mu_m}{}_{\nu_1 \ldots \nu_n} = \Lambda^{\mu_1}{}_{\rho_1} \ldots \Lambda^{\mu_m}{}_{\rho_m} \Lambda^{\lambda_1}{}_{\nu_1} \ldots \Lambda^{\lambda_n}{}_{\nu_n} T^{\rho_1 \ldots \rho_m}{}_{\lambda_1 \ldots \lambda_n}. \tag{A.8}$$

Contracción con la métrica $\eta_{\mu\nu}$ cambia el rango de un tensor, por ejemplo:

$$\eta_{\lambda\sigma} T^{\mu\nu\rho\lambda} = T^{\mu\nu\rho}{}_\sigma, \qquad\qquad \eta_{\rho\lambda} T^{\mu\nu\rho\lambda} = T^{\mu\nu\rho}{}_\rho = T^{\mu\nu}. \tag{A.9}$$

A.2. Geometría diferencial

Bajo un cambio general de coordenadas $y^\alpha = y^\alpha(x^\mu)$, un tensor de rango (m, n) transforma como

$$T^{\mu_1 \ldots \mu_m}{}_{\nu_1 \ldots \nu_n} = \frac{\partial x^{\mu_1}}{\partial y^{\alpha_1}} \ldots \frac{\partial x^{\mu_m}}{\partial y^{\alpha_m}} \frac{\partial y^{\beta_1}}{\partial x^{\nu_1}} \ldots \frac{\partial y^{\beta_n}}{\partial x^{\nu_n}} \, T^{\alpha_1 \ldots \alpha_m}{}_{\beta_1 \ldots \beta_n} \tag{A.10}$$

Contracción con la métrica $g_{\mu\nu}$ cambia el rango de un tensor, por ejemplo:

$$g_{\lambda\sigma} T^{\mu\nu\rho\lambda} = T^{\mu\nu\rho}{}_\sigma, \qquad\qquad g_{\rho\lambda} T^{\mu\nu\rho\lambda} = T^{\mu\nu\rho}{}_\rho = T^{\mu\nu}. \tag{A.11}$$

La derivada covariante de un vector contravariante V^ν y de un vector covariante V_ν viene dada por

$$\nabla_\mu V^\nu = \partial_\mu V^\nu + \Gamma^\nu_{\mu\rho} V^\rho, \qquad\qquad \nabla_\mu V_\nu = \partial_\mu V_\nu - \Gamma^\rho_{\mu\nu} V_\rho, \tag{A.12}$$

y de un tensor de rango (m, n) por

$$\begin{aligned} \nabla_\rho T^{\mu_1 \ldots \mu_m}{}_{\nu_1 \ldots \nu_n} &= \partial_\rho T^{\mu_1 \ldots \mu_m}{}_{\nu_1 \ldots \nu_n} + \Gamma^{\mu_1}_{\rho\lambda} T^{\lambda\mu_2 \ldots \mu_m}{}_{\nu_1 \ldots \nu_n} + \ldots \qquad (A.13) \\ &+ \Gamma^{\mu_m}_{\rho\lambda} T^{\mu_1 \ldots \mu_{m-1}\lambda}{}_{\nu_1 \ldots \nu_n} - \Gamma^\lambda_{\rho\nu_1} T^{\mu_1 \ldots \mu_m}{}_{\lambda\nu_2 \ldots \nu_n} - \ \ldots \ - \Gamma^\lambda_{\rho\nu_n} T^{\mu_1 \ldots \mu_m}{}_{\nu_1 \ldots \nu_{n-1}\lambda}. \end{aligned}$$

La conexión transforma bajo un cambio general de coordenadas $y^\alpha = y^\alpha(x^\mu)$ como

$$\Gamma^\gamma_{\alpha\beta} = \frac{\partial x^\mu}{\partial y^\alpha} \frac{\partial x^\nu}{\partial y^\beta} \frac{\partial y^\gamma}{\partial x^\rho} \Gamma^\rho_{\mu\nu} - \frac{\partial^2 y^\gamma}{\partial x^\mu \partial x^\nu} \frac{\partial x^\mu}{\partial y^\alpha} \frac{\partial x^\nu}{\partial y^\beta}, \tag{A.14}$$

de modo que la derivada covariante de un tensor de rango (m, n) es un tensor de rango $(m, n+1)$.

Un tensor $T^{\mu_1 \ldots \mu_m}{}_{\nu_1 \ldots \nu_n}$ es transportado paralelo a lo largo de la curva C, si la derivada covariante del tensor a lo largo de la curva es cero,

$$u^\rho \nabla_\rho T^{\mu_1 \ldots \mu_m}{}_{\nu_1 \ldots \nu_n} = 0, \tag{A.15}$$

donde u^μ es el vector tangente a la curva C.

Un variedad tiene una curvatura intrínsica si el transporte paralelo de vectores y tensores entre dos puntos depende de la curva entre estos dos puntos. Concretamente

$$\begin{aligned} [\nabla_\mu, \nabla_\nu] S^{\rho_1 \ldots \rho_m}{}_{\lambda_1 \ldots \lambda_n} &= R_{\mu\nu\sigma}{}^{\rho_1} S^{\sigma\rho_2 \ldots \rho_m}{}_{\lambda_1 \ldots \lambda_n} + \ \ldots \ + R_{\mu\nu\sigma}{}^{\rho_m} S^{\rho_1 \ldots \rho_{m-1}\sigma}{}_{\lambda_1 \ldots \lambda_n} \qquad (A.16) \\ -R_{\mu\nu\lambda_1}{}^\sigma S^{\rho_1 \ldots \rho_m}{}_{\sigma\lambda_2 \ldots \lambda_n} &+ \ \ldots \ - R_{\mu\nu\lambda_n}{}^\sigma S^{\rho_1 \ldots \rho_m}{}_{\lambda_1 \ldots \lambda_{n-1}\sigma} - T^\sigma_{\mu\nu} \nabla_\sigma S^{\rho_1 \ldots \rho_m}{}_{\lambda_1 \ldots \lambda_n}, \end{aligned}$$

donde el tensor de Riemann $R_{\mu\nu\rho}{}^{\lambda}$ y el tensor de torsión $T^{\rho}_{\mu\nu}$ vienen dados por

$$R_{\mu\nu\rho}{}^{\lambda} = \partial_\mu \Gamma^\lambda_{\nu\rho} - \partial_\nu \Gamma^\lambda_{\mu\rho} + \Gamma^\lambda_{\mu\sigma}\Gamma^\sigma_{\nu\rho} - \Gamma^\lambda_{\nu\sigma}\Gamma^\sigma_{\mu\rho}, \qquad T^\rho_{\mu\nu} = \Gamma^\rho_{\mu\nu} - \Gamma^\rho_{\nu\mu}. \tag{A.17}$$

Las expresiones para la geodésica afín y la geodésica métrica son respectivamente

$$\ddot{x}^\sigma + \Gamma^\sigma_{\nu\rho}\dot{x}^\nu\dot{x}^\rho = 0, \qquad \ddot{x}^\sigma + \left\{ \begin{array}{c} \lambda \\ \nu\rho \end{array} \right\} \dot{x}^\nu\dot{x}^\rho = 0, \tag{A.18}$$

donde los símbolos de Christoffel vienen dados por

$$\left\{ \begin{array}{c} \lambda \\ \nu\rho \end{array} \right\} = \frac{1}{2} g^{\lambda\sigma}\Big(\partial_\nu g_{\sigma\rho} + \partial_\rho g_{\nu\sigma} - \partial_\sigma g_{\nu\rho}\Big). \tag{A.19}$$

La expresión de la geodésica afín sale pidiendo que el vector tangente a la curva sea transportado paralelamente a lo largo de la curva, $u^\nu \nabla_\nu u^\mu = 0$, mientras la ecuación de la geodésica métrica sale extremizando el funcional $s = \int d\tau \sqrt{g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu}$. En espacios lorentzianos hay que especificar el tipo de geodésica:

$$g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu = \varepsilon, \qquad \text{donde} \qquad \varepsilon = \begin{cases} 1 & \text{si } x^\mu(\tau) \text{ es temporal,} \\ 0 & \text{si } x^\mu(\tau) \text{ es nulo,} \\ -1 & \text{si } x^\mu(\tau) \text{ es espacial.} \end{cases} \tag{A.20}$$

Si la conexión $\Gamma^\rho_{\mu\nu}$ satisface las siguientes dos condiciones,

1. La conexión es simétrica: $\Gamma^\rho_{\mu\nu} = \Gamma^\rho_{\nu\mu}$,
2. La derivada covariante de la métrica es cero: $\nabla_\mu g_{\nu\rho} = 0$,

entonces $\Gamma^\rho_{\mu\nu}$ es la conexión de Levi-Civita y está determinada por completo por la métrica y coincide con los símbolos de Christoffel

$$\Gamma^\lambda_{\rho\mu} = \left\{ \begin{array}{c} \lambda \\ \rho\mu \end{array} \right\} = \frac{1}{2} g^{\lambda\sigma}\Big(\partial_\rho g_{\sigma\mu} + \partial_\mu g_{\rho\sigma} - \partial_\sigma g_{\rho\mu}\Big). \tag{A.21}$$

De esta expresión se puede derivar la identidad útil,

$$\Gamma^\nu_{\mu\nu} = \frac{1}{\sqrt{|g|}}\partial_\mu \sqrt{|g|}, \tag{A.22}$$

donde hemos utilizado que las siguientes identidades para el determinante g de la métrica

$$\delta g^{\mu\nu} = g^{\mu\rho} g^{\nu\lambda}\delta g_{\rho\lambda}, \qquad \delta g = g\, g^{\mu\nu}\delta g_{\mu\nu} = -g\, g_{\mu\nu}\delta g^{\mu\nu},$$

$$\delta\sqrt{|g|} = \frac{1}{2}\sqrt{|g|}\, g^{\mu\nu}\delta g_{\mu\nu} = -\frac{1}{2}\sqrt{|g|}\, g_{\mu\nu}\delta g^{\mu\nu}. \tag{A.23}$$

El tensor de Riemann tiene las siguientes propiedades (asumiendo la conexión de Levi-Civita):

$$R_{\mu\nu\rho\lambda} = -R_{\nu\mu\rho\lambda}, \qquad R_{\mu\nu\rho\lambda} = -R_{\mu\nu\lambda\rho}, \qquad R_{\mu\nu\rho\lambda} = R_{\rho\lambda\mu\nu}, \tag{A.24}$$

$$R_{\mu\nu\rho\lambda} + R_{\mu\rho\lambda\nu} + R_{\mu\lambda\nu\rho} = 0, \qquad \nabla_\mu R_{\nu\rho\lambda\sigma} + \nabla_\nu R_{\rho\mu\lambda\sigma} + \nabla_\rho R_{\mu\nu\lambda\sigma} = 0.$$

A través de contracciones con la métrica se puede construir el tensor de Ricci $R_{\mu\nu}$, el escalar de Ricci R y el tensor de Einstein $G_{\mu\nu}$

$$\begin{aligned} R_{\mu\nu} &= R_{\mu\rho\nu}{}^{\rho} = g^{\rho\lambda}R_{\mu\rho\nu\lambda} = \partial_\mu \Gamma^\lambda_{\nu\lambda} - \partial_\lambda \Gamma^\lambda_{\mu\nu} + \Gamma^\lambda_{\mu\sigma}\Gamma^\sigma_{\lambda\nu} - \Gamma^\lambda_{\mu\nu}\Gamma^\sigma_{\lambda\sigma}, \\ R &= g^{\mu\nu}R_{\mu\nu} \qquad G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R. \end{aligned} \tag{A.25}$$

El tensor de Ricci, el escalar de Ricci y el tensor de Einstein satisfacen las seguientes propiedades (asumiendo la conexión de Levi-Civita):

$$\begin{aligned} R_{\mu\nu} &= R_{\nu\mu}, \qquad \nabla_\mu R_\nu{}^\mu = \frac{1}{2}\partial_\nu R, \\ G_{\mu\nu} &= G_{\nu\mu}, \qquad \nabla_\mu G_\nu{}^\mu = 0. \end{aligned} \tag{A.26}$$

A.3. Relatividad general

Las ecuaciones de Einstein son

$$R_{\mu\nu} \ - \ \frac{1}{2}g_{\mu\nu}R \ = \ -8\pi G_N T_{\mu\nu}, \tag{A.27}$$

o, en la forma sin traza,

$$R_{\mu\nu} = -8\pi G_N\Big(T_{\mu\nu} - \frac{1}{2}g_{\mu\nu}T\Big). \tag{A.28}$$

donde la expresión para el tensor de energía-momento viene dada por

$$\begin{array}{ll} T^{\mu\nu} = \rho_0 \, u^\mu u^\nu & \text{para polvo,} \\ T^{\mu\nu} = (\rho_0 + P)u^\mu u^\nu - P g^{\mu\nu} & \text{para un fluido perfecto,} \\ T^{\mu\nu} = \ -F^{\mu\rho}F^\nu{}_\rho \ + \ \frac{1}{4}g^{\mu\nu}F_{\rho\lambda}F^{\rho\lambda} & \text{para electromagnetismo,} \\ T^{\mu\nu} = \Lambda g^{\mu\nu} & \text{una constante cosmológico.} \end{array} \tag{A.29}$$

Las ecuaciones de Einstein se pueden obtener de la acción

$$S = \int d^4x\sqrt{|g|}\Big[\frac{1}{2\kappa}R + \mathcal{L}_{\rm mat}\Big], \tag{A.30}$$

donde $\kappa = 8\pi G_N$ y $\mathcal{L}_{\rm mat}$ es el lagrangiano que describe la dinámica y las interacciones de los campos no-gravitacionales. El tensor de energía-momento $T_{\mu\nu}$ correspondiente a estos campos no-gravitacionales se puede obtener a través de una variación del lagrangiano $\mathcal{L}_{\rm mat}$:

$$T_{\mu\nu} = \frac{2}{\sqrt{|g|}}\frac{\delta(\sqrt{|g|}\mathcal{L}_{\rm mat})}{\delta g^{\mu\nu}}. \tag{A.31}$$

La ley de conservación de energía y momento implica que $\nabla_\mu T^{\mu\nu} = 0$.

Una ecuación útil para derivar las ecuaciones de Einstein desde una acción es la identidad de Palatini:

$$\delta R_{\mu\nu} \ = \ \nabla_\mu(\delta\Gamma^\lambda_{\lambda\nu}) - \nabla_\lambda(\delta\Gamma^\lambda_{\mu\nu}) + T^\sigma_{\mu\lambda}(\delta\Gamma^\lambda_{\sigma\nu}), \tag{A.32}$$

donde $\Gamma^\rho_{\mu\nu}$ es una conexión arbitraria y $T^\rho_{\mu\nu}$ el tensor de torsión.

La segunda ley de Newton y las leyes de Maxwell en espacios curvos vienen dadas por

$$\begin{array}{c} m_0\Big(\ddot{x}^\mu + \Gamma^\mu_{\nu\rho}\dot{x}^\nu\dot{x}^\rho\Big) \ = \ f^\mu, \\ \nabla_\mu F^{\mu\nu} = j^\nu, \qquad \partial_\mu F_{\nu\rho} + \partial_\rho F_{\mu\nu} + \partial_\nu F_{\rho\mu} = 0, \end{array} \tag{A.33}$$

donde $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. La ecuación inhomogénea se puede rescribir, utilizando (A.22) como

$$\partial_\mu\Big[\sqrt{|g|}\,F^{\mu\nu}\Big] = \sqrt{|g|}\,j^\nu. \tag{A.34}$$

La solución de Schwarzschild, en las coordenadas de Schwarzschild, Eddington-Finkelstein y Kruskal viene dada respectivamente por

$$\begin{aligned} ds^2 \ &= \ \Big(1 - \frac{2M}{r}\Big)dt^2 \ - \ \Big(1 - \frac{2M}{r}\Big)^{-1}dr^2 \ - \ r^2\Big(d\theta^2 + \ \sin^2\theta d\varphi^2\Big) \\ &= \ \Big(1 - \frac{2M}{r}\Big)d\tilde{t}^2 \ - \ \frac{4M}{r}d\tilde{t}dr \ - \ \Big(1 + \frac{2M}{r}\Big)dr^2 \ + \ r^2\Big(d\theta^2 + \ \sin^2\theta d\varphi^2\Big) \\ &= \ \frac{16M^2}{r}e^{-r/2M}\Big(dT^2 - dR^2\Big) - r^2 d\Omega_2^2, \end{aligned} \tag{A.35}$$

donde

$$\tilde{t} \;=\; t \;+\; 2M\log(r-2M),$$

$$T = e^{r/4M}\sqrt{r-2M}\;\sinh\Big(\frac{t}{4M}\Big), \qquad R = e^{r/4M}\sqrt{r-2M}\,\cosh\Big(\frac{t}{4M}\Big). \qquad \text{(A.36)}$$

Apéndice B

Bibliografía

Hay numerosos libros sobre la Teoría de la Relatividad, a muchos niveles distintos y cada uno con sus propios puntos de vistas y énfasis particulares.

En la preparación de este curso me he dejado inspirar por las lineas de razonamiento de tres libros excelentes:

- Ray d'Inverno, *Introducing Einstein's Relativity*, Oxford University Press, 1992.
- Sean M. Carroll, *Spacetime and Geometry*, Addison-Wesley, 2004.
- Derek F. Lawden, *An Introduction to Tensor Calculus and Relativity*, Wiley, 1962.

El libro de d'Inverno es pedagógicamente un libro muy bueno, ideal para aprender Relatividad, tanto especial como general. Aparte de su buena pedagogía, cubre temas muy amplios y de manera bastante profunda para un libro de introducción al tema.
El libro de Carroll es igual de bueno que el de d'Inverno explicando, pero tiene un estilo un poco más teórico. Presta poca atención a tests experimentales pero hace bastante referencia a temas modernas como cosmología, teoría cuántica de campos o temas de teoría de cuerdas.
Finalmente el libro de Lawden me ayudó mucho cuando yo empecé a estudiar Relatividad. Tiene una excelente introducción de álgebra de tensores y una buenísima discusión sobre las invariancias de la mecánica newtoniana, la teoría de Maxwell, el Principio de la Relatividad y el Principio de Equivalencia.

Aparte de estos tres libros hay muchos que merecen la pena. Mencionaré aquí algunos libros clásicos, sin comentarlos:

- J.B. Hartle, *Gravity: An introduction to Einstein's General Relativity*, Addison-Wesly, 2002.
- J. Foster and J.D. Nightingale, *A Short course in General Relativity*, Springer-Verlag, 1995.
- B.F. Schutz, *A First Course in General Relativity*, Cambridge University Press, 1985.
- S. Weinberg, *Gravitation and Cosmology*, Wiley, 1972.
- C. Misner, K. Thorne, A. Wheeler, *Gravitation*, Freeman, 1973.
- R. Wald, *General Relativity*, Chicago University Press, 1984.

En internet también se puede encontrar numerosos cursos, notas y review sobre Relatividad:

- S. Carroll, *Lecture Notes on General Relativity*, [1]
`http://fr.arXiv.org/abs/gr-qc/9712019`
- M. Blau, *Lecture Notes on General Relativity*,
`http://www.unine.ch/phys/string/lecturesGR.ps.gz`
- G. 't Hooft, *Introduction to General Relativity*,
`http://www.phys.uu.nl/ thooft/lectures/genrel.ps`
- Stefan Waner, *Introduction to Differential Geometry and General Relativity*,
`http://people.hofstra.edu/faculty/Stefan_Waner/RealWorld/pdfs/DiffGeom.pdf`

[1]Estas notas son en realidad un primera versión del libro mencionado arriba. El libro es un poquito más extenso, pero en grandes lineas las notas son casi iguales.

Apéndice C

Referencias

Artículos y libros de texto

- J.S. Bell, *Speakable and unspeakable in quantum mechanics*, Cambridge University Press, 1987.
- J.D. Bjorken and S.D. Drell, *Relativistic Quantum Mechanics*, McGraw-Hill (1964).
- M. Blau, *Lecture Notes on General Relativity*, http://www.blau.itp.unibe.ch/lecturesGR.pdf.
- M. Blau, *Lecture Notes on Plane Waves and Penrose Limits*, http://www.blau.itp.unibe.ch/lecturesPP.pd
- A. Bohm, *Quantum Mechanics*, Springer-Verlag, 1979.
- R. Bousso, *The holographic Principle*, http://arXiv.org/abs/hep-th/0203101.
- S. M. Carroll, *Spacetime and Geometry*, Addison-Wesley, 2004.
- S. Carroll, *Lecture Notes on General Relativity*, http://arXiv.org/abs/gr-qc/9712019.
- T.M. Davis, C.H. Lineweaver, *Expanding Confusion: common misconceptions of cosmological horizons and the superluminal expansion of the Universe*, http://arxiv.org/abs/astro-ph/0310808.
- Victor de Cauter *Van punt tot Einsteins gravitatiewet: de algemene relativiteitstheorie in een notedop*, Volkssterrenwacht Urania, 1988.
- N. de Kort (red.), *Klassieke Mechanica: Elektriciteit en magnetisme*, Stiching Teleac, 1990.
- R. d'Inverno, *Introducing Einstein's Relativity*, Oxford University Press, 1992.
- A. Einstein, *Relativity, the special and general theory*, Methuen & Co, 1920.
- A. Einstein, F.A. Davis, *The Principle of Relativity*, Dover Publishing Company, 1952.
- J.P. Elliott and P.G. Dawber, *Symmetry in Physics, Volume 1 & 2*, Macmillan, 1979.
- R. Feynman, R. Leighton, and M. Sands, *The Feynman Lectures on Physics: Volume II*, Addison-Wesley, 1964.
- R. P. Feynman, F. B. Morinigo, W. G. Wagner, D. Pines, *The Feynman Lectures on Gravitation*, Addison-Wesley, 1995.

- V.A. Fock *Fundamentals of Quantum Mechanics*, URSS Publishers, 1978.
- M. Forger and M. Römer, Annals Phys. 309 (2004) 306-389, http://arxiv.org/abs/hep-th/0307199
- J. Foster and J.D. Nightingale, *A Short course in General Relativity*, Springer-Verlag, 1995.
- L. Garay, *TAE 2008: General relativity and gravitation*, http://lattice.ft.uam.es/TAE2008/lectures_LJG.pdf.
- H. Goldstein, *Classical Mechanics*, Addison-Wesley, 1980.
- S. Hawking and Ellis, *The large scale structure of space-time*, Cambridge University Press, 1973.
- J.I. Illana, *El modelo estándar y su fenomenología. Parte 1*, http://www.ugr.es/local/jillana/curso.pdf
- D.D. Ivanenko and G.A. Sardanashvili, *Gravitación*, URSS, 2005.
- C. Itzykson and J.B Zuber, *Quantum field theory*, McGraw-Hill, 1985.
- J.D. Jackson, *Classical Electromagnetism*, John Wiley & Sons, 1962.
- L.D. Landau and E.M. Lifshitz, *Mechanics and Electrodynamics*, Pergamon Press, 1972.
- D.F. Lawden, *An Introduction to Tensor Calculus and Relativity*, Wiley, 1962.
- M. Maggiore, *A Modern Introduction to Quantum Field Theory*, Oxford University Press, 2005.
- M. Maggiore, *Gravitational waves, Volume 1, Theory and Experiments*, Oxford University Press, 2007.
- T.M. Marsh, *Notes for PH 421, Cosmology*.
- C. Misner, K. Thorne, J.A. Wheeler, *Gravitation*, Freeman, 1973.
- S. Montero Modino, *The Lorentz abd Poincaré Groups*.
- U. Moschella, *The de Sitter and anti-de Sitter Sightseeing Tour*, Séminaire Poincaré 1 (2005) 1-12.
- M. Nakahara, *Geometry, Topology and Physics*, Institute of Physics Publishing, 1990.
- T. Ortín *Gravity and Strings*, Cambridge University Press, 2004.
- M.E. Peskin and D.V. Schroeder, *An introduction to Quantum Field Theory*, Addison-Wesley, 1995.
- P. Ramond, *Field Theory: A Modern Primer*, Westview Press, 1997.
- M. Robinson, K. Bland, G. Cleaver, J. Dittmann, *A Simple Introduction to Particle Physics*, http://arxiv.org/abs/0810.3328.
- M.B. Robinson, T. Ali, G.B. Cleaver, *A Simple Introduction to Particle Physics Part II*, http://arxiv.org/abs/0908.1395
- L.H. Ryder, *Quantum Field Theory*, Cambridge University Press, 1985.
- B.F. Schutz, *A First Course in General Relativity*, Cambridge University Press, 1985.
- H. Stefani, *General Relativity*, Cambridge University Press, 1982.

- W.O. Straub, *Weyl spinors and Dirac's electron theory*, http://www.weylmann.com/weyldirac.pdf.
- G. 't Hooft, *Introduction to General Relativity*, http://www.phys.uu.nl/∼thooft/lectures/genrel.ps.
- P.K. Townsend, *Black holes*, http://arxiv.org/abs/gr-qc/9707012.
- M. Trodden and S.M. Carroll, *TASI Lectures: Introduction to Cosmology*, http://arxiv.org/abs/astro-ph/0401547.
- P. van Kampen, *Lorentz contraction and current-carrying wires*, Eur. J. Phys. 29 (2008) 879.
- T. van Proeyen, *Tools for supersymmetry*, http://arxiv.org/abs/hep-th/9910030.
- T. Van Riet, *Cosmic acceleration in Kaluza-Klein supergravity*, Ph-D thesis Rijksuniversiteit Groningen, 2007.
- M.J.G. Veltman, B.Q.P.J. de Wit and G. 't Hooft, *Lie-groepen in de Fysica*, http://www.phys.uu.nl/∼thooft/lectures/lieg03.pdf
- M. Visser, *The Kerr spacetime: A brief introduction*, http://arxiv.org/abs/0706.0622.
- R. Wald, *General Relativity*, Chicago University Press, 1984.
- S. Weinberg, *Gravitation and Cosmology*, Wiley, 1972.

Libros divulgativos y de historia

- J. Daintith and D. Gjertsen (ed.), *A dictionary of scientists*, Oxford University Press, 1999.
- D. Dieks, *Space, Time and Coordinates in a Rotating World*, http://arxiv.org/abs/1002.0130.
- F.G. Droste, *Simon Stevin*, Uitgeverij Aspekt, 2007.
- D.W. Ebner, *How Hilbert has found the Einstein equations before Einstein and forgeries of Hilbert's page proofs*, http://fr.arxiv.org/abs/physics/0610154.
- A. Einstein and L. Infeld, *Evolution of Physics*, Cambridge University Press, 1938.
- P. Gallisson, *Einstein's clocks and Poincaré's maps*, Sceptre, 2003.
- S. Hacyan, *Geometry as an object of experience: Kant and the missed debate between Poincaré and Einstein*, http://arxiv.org/abs/0712.2222.
- E.R. Harrison, *Cosmology, the science of the Universe*, Cambridge University Press, 1981.
- S. Hawking (ed.), *On the shoulders of Giants*, Running Press, 2002.
- D.R. Hofstadter, *Gödel, Escher, Bach, an Eternal Golden Braid*, Penguin Books, 1979.
- J.D. Jackson and L.B. Okun, *Historical roots of gauge invariance*, Rev.Mod.Phys., Vol 73, No 3, 663-680.
- D. Kennefick, *Controversies in the History of the Radiation Reaction problem in General Relativity*, http://arxiv.org/abs/gr-qc/9704002.
- H. Kragh, *Quantum generations*, Princeton University Press, 1999.

- E. Mach, *Natuurkunde, Wetenschap en Filosofie*, Boom, 1980.
- A.I. Miller, *Empire of the Stars*, Houghton Mifflin, 2005.
- J.D. Norton, *Einstein's Investigations of Galilean Covariant Electrodynamics Prior to 1905*, Arch.Hist.Sci. 59 (2004) 45-105.
- H. Nussbaumer and L. Bieri, *Discovering the Expanding Universe*, Cambridge University Press, 2009.
- R. Penrose, *The Road to Reality: A Complete Guide to the Laws of the Universe*, Vintage Books, 2004.
- H. Reichenbach, *The philosophy of Space and Time*, Dover Publications Inc., 1957.
- B. Russell, *History of Western Philosophy*, Simon & Schuster Inc., 1972.
- E.L. Schücking, *Einstein's Apple and Relativity's Gravitational Field*, `http://arxiv.org/abs/0903.3768`.
- D. Sobel, *Galileo's Daughter*, Fourth Estate, 1999.
- T. Standage, *The Neptune file*, Penguin Books, 2000.
- John Waller, *Einstein's luck - The truth behind some of the greatest scientific discoveries*, Oxford university press, 2002